STATISTICAL CONSIDERATIONS IN THE DESIGN AND ANALYSIS OF CROSS-OVER TRIALS

by

GILBERT HENEAGE MORREY

THESIS SUBMITTED FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY
TO THE COUNCIL FOR NATIONAL
ACADEMIC AWARDS

DEPARTMENT OF MATHEMATICAL SCIENCES
LEICESTER POLYTECHNIC

August 1991

Statistical Considerations in the Design

and Analysis of Cross-over Trials

by

Gilbert Heneage Morrey

ABSTRACT

Cross-over designs are frequently used in clinical trials, usually to compare two treatments. It has long been known that the simple two-period cross-over is adversely affected by the presence of carry-over effects, but it is nevertheless still commonly used. This thesis examines the problems caused by carry-over effects in detail and considers ways of overcoming these problems, either by using a more complex cross-over design, or by carrying out a more sophisticated statistical analysis.

Only designs for comparing two treatments are considered, but analysis for continuous and binary variables is covered.

ACKNOWLEDGEMENTS

I wish to thank Professer Derek Teather for his advice and support throughout, and Professer Peter Freeman for his advice in the early stages. I am indebted to Dr. Allan Ebbutt for his permission to use his data, and to the staff of the James Went computing node for their advice on computing matters and NAG routines.

CONTENTS

	Page
Chapter 1: Introduction	
1.1 Experimental Design Considerations	1
1.2 Application of Cross-over Designs	2
1.3 Clinical Trials	3
1.4 Carry-over	4
1.5 The Two-period Cross-over	5
1.6 Multi-period Cross-overs	6
1.7 Analysis Considerations	7
1.8 Discussion	8
Chapter 2: Carry-over	
2.1 Introduction	9
2.2 The Two-period Cross-over	10
2.3 Effect of Carry-over on the Model	12
2.4 Testing for Carry-over	15
2.5 The Nature of Carry-over Effects	16
Chapter 3: Higher-order Cross-overs	
3.1 Introduction	21
3.2 Increasing the Number of Sequences	21
3.3 Increasing the Number of Treatment Periods	
3.3.1 Introduction	23
3.3.2 Efficiency of Three-period Designs	26
3.3.3 Properties of Efficient Designs	27
3.3.4 Optimal Designs	31

3.4 Designs to Consider Further 33

Chapter 4: Classical Analysis for Continuous Response Variables: Two-period Designs

4.1 Introduction 35

4.2 The Simple Two-period Cross-over 36

4.3 The "Complete" Two-period Cross-over 45

4.3.1 Baseline Observations 49

4.3.2 Washout Observations 52

4.4 Review and Preview 53

Chapter 5: Classical Analysis for Continuous Response Variables: Three-period Designs

5.1 Introduction 55

5.2 The Two Sequence Design Without Baselines 55

5.2.1 The Two Sequence Design with Baselines 63

5.2.2 Baseline and Washout Observations 64

5.3 Four Sequence Designs 68

5.4 Review 73

Chapter 6: The Two-period Cross-over with Binary Data

6.1 Introduction 74

6.2 Models for Binary Data 74

6.3 The Mainland-Gart Test 79

6.4 Zimmerman & Rahlf's Model 84

6.5 Cox & Plackett's Models 85

6.6 Fidler's Model 87

6.7 Kenward & Jones's Model 89

6.8 Other Tests

6.8.1 Prescott's Test 92

6.8.2 Tests of Carry-over 93

6.9 Discussion 95

Chapter 7: Higher-order Cross-overs with Binary Data

7.1 Introduction 96

7.2 The Complete Two-period Cross-over 96

7.3 Three-period Designs 104

7.3.1 The Two-sequence Design 105

7.3.2 The Four-sequence Design 113

7.4 Discussion 116

Chapter 8: Power of Tests

8.1 Introduction 117

8.2 Calculation of Multivariate Normal Probabilities 120

8.3 The Parallel Design 120

8.4 The Simple Two-period Cross-over 122

8.5 The "Complete" Two-period Cross-over 136

8.6 The Three-period Designs 138

8.6.1 The Two-sequence Design 140

8.6.2 The Four-sequence Design 145

8.7 Discussion 148

Chapter 9: Bayesian Analysis of Cross-over Designs

9.1 Introduction 150

9.2 Results for the Simple Two-period Cross-over 150

9.3 The Two-period Design with Baselines 153

9.4 Gibbs Sampling 157

9.5 A Simple Example 160

9.6 Gibbs Sampling for the Two-period Cross-over 167

9.7 Adding Carry-over to the Model 170

9.8 Three-period Designs 171

9.8.1 Results 172

9.9 Effectiveness of Gibbs Sampling 178

9.10 Discussion 180

Chapter 10: Discussion

10.1 Introduction 182

10.2 The Problems of Carry-over 182

10.3 Alternative Designs 183

10.4 Alternative Analyses 185

10.5 Cross-over Designs and Clinical Trials 186

10.6 Concluding Remarks 188

References 189

Appendices

Appendix A Calculation of Multivariate Normal Probabilities

Appendix B Program to Calculate the Power for the Two-period Cross-over

Appendix C Program to Calculate the Power for the Three-period Cross-over

Appendix D Derivation of Posterior Distributions for the Two-period Cross-over

Appendix E Derivation of Posterior Distributions for the Three-period Two-sequence Cross-over

Appendix F Data Sets

LIST OF FIGURES

		Page
Fig 1	Power of the test for treatment difference for the parallel design with baseline observations	123
Fig 2	Performance of the test procedure for the two-period cross-over, N = 20	131
Fig 3	Performance of the test procedure for the two-period cross-over, N = 40	132
Fig 4	Performance of the test procedure for the two-period cross-over, N = 60	133
Fig 5	Power of the test for treatment difference for the "complete" two-period cross-over	139
Fig 6	Sequence of tests for the three-period two-sequence design	142
Fig 7	Performance of the test procedure for the three-period cross-over	146
Fig 8	Power of the test for treatment difference for the four-sequence three-period cross-over	147

Fig 9 Sequence of operations in Gibbs sampling 159

Fig 10 Gibbs sampler for the simple example 161

Fig 11 Convergence of the Gibbs sampler for the simple example Cycles 1 & 2 163

Fig 12 Convergence of the Gibbs sampler for the simple example Cycles 12 & 13 164

Fig 13 Convergence of the Gibbs sampler for the simple example Cycles 24 & 25 165

Fig 14 Estimated and theoretical posteriors for θ_1 166

Fig 15 Gibbs sampler for the two-period cross-over 170

Fig 16 Estimated posterior for the enuresis data. τ, given $\alpha = 0$; τ & α 171a

Fig 17 Estimated posterior for τ & α, given $\beta = 0$. Initial estimates $\tau = 2.96$, $\alpha = 1.336$ 174

Fig 18 Estimated posterior for τ, α & β.
Initial estimates τ = 2.96, α = 1.336,
β = 8.116 175

Fig 19 Estimated posterior for τ & α, given β = 0.
Initial estimates τ = 2.246, α = -0.094 176

Fig 20 Estimated posterior for τ, α & β.
Initial estimates τ = 2.246, α = -0.094,
β = -2.86 177

LIST OF TABLES

		Page
Table 1	Leading factors of the variance of unbiased estimators of contrasts between treatment and carry-over effects for various two-treatment cross-over designs	25
Table 2	Variances of estimators for the two-period cross-over design	41
Table 3	Sequences in the three designs	69
Table 4	Implication of patterns in the three-period design	106
Table 5	Performance of the test procedure for the three-period cross-over N = 20	125
Table 6	Performance of the test procedure for the three-period cross-over N = 40	127
Table 7	Performance of the test procedure for the three-period cross-over N = 60	129

Chapter 1: Introduction

1.1 Experimental Design Considerations

Science has progressed by proposing hypotheses which are then tested in controlled experiments. If a scientist wishes to demonstrate that a new procedure or treatment has a beneficial effect, an experiment to compare the new treatment with other possible treatments will be necessary. The simplest such experiment would involve the application of the new treatment to a set of experimental units, and the comparison of the outcomes for these units with those for a set of "controls". These "controls" would be another set of experimental units which are subjected to exactly the same conditions as the treated units except that they do not receive the new treatment. Thus the word "controls" in this context implies that differences other than the treatment under test are not allowed to affect the experiment. Clearly, a possible difference which could affect the experiment is variability in the experimental units themselves, and a desire to control this has lead to the common-sense idea of grouping the experimental units in blocks or pairs, within which they are as similar as possible. Although this works well in many cases, in some circumstances, particularly where the experimental units are animals or people, the process of identifying similar individuals is very difficult, or impossible. For example, in the study

of post-operative pain, not only is there great variation between individuals in the level of pain experienced, but experimenters have been unable to correlate the pain with other factors such as age, sex, disease etc., and so unable to predict the level of pain in order to match individuals. (Keats et al. [1950]) For such cases, the idea of obtaining the equivalent of a paired comparison, by using "the subject as his own control" was devised, wherein the test and control treatments alternate in successive treatment periods, using the same experimental unit repeatedly. An experiment which uses a succession of treatments on each experimental unit in this way is called a change-over or cross-over design.

1.2 Application of Cross-over Designs

The purpose of matching or pairing is to increase the precision of the experiment by reducing the variability of the experimental units that are compared. Since an alternative way of increasing precision is to increase the number of experimental units in the treated and control groups, an alternative way of looking at the benefit of matching is that the same precision may be obtained with fewer observations. Cross-over designs have an immediate intuitive appeal because, by applying the treatments to the same experimental unit, it appears that we are ensuring that the experimental units are matched as exactly as possible, and that the potential benefits

of matching will be maximised. However, it is clear that this will only be the case if the treatments do not effect any permanent change to the experimental unit, so that, when subsequent treatments are applied, the experimental unit is in the same state as when the first treatment was applied. An example of this, and an early application of cross-over trials, is the effect of diet on milk yield in dairy cows. (Cochran et al.[1941]) A diet may be expected to influence the milk yield, but would not be expected to cause any lasting physiological change to the cow. On the other hand, there may be external factors, such as the variation in milk yield in the cow's lactation cycle, or the natural diminution of pain with the passage of time in the case of post-operative pain, which could affect the result of the experiment. In order to control factors such as these, a balanced design must be used so that any changes will affect each treatment equally. This requirement meant that early cross-over designs were based on latin squares with rows as treatment periods and columns as experimental units.

1.3 Clinical Trials

Although cross-over designs have been used in many fields of application, they are perhaps most commonly used in medical experiments, when comparing palliative treatments. One reason for this is the large variability

between subjects, and the difficulty in predicting how a patient will react to a treatment. Another reason is that it has often proved easier to relieve a symptom than to understand the underlying cause and cure the disease e.g. pain in arthritis, blood pressure in hypertension. Because of the importance of cross-over designs to clinical trials, this thesis will concentrate on this area of application, although the majority of the results would apply to cross-over designs generally.

1.4 Carry-over

Even if the treatments cause no lasting change to the experimental units, it is possible that their effect will not be confined to the treatment period in which they are applied. Thus a treatment may have some residual effect which carries over to subsequent treatment periods. If this happens then the apparent effect of subsequent treatments will be a combination of the direct effect of that treatment and any carry-over from previous treatments. Since it is likely that, in any future application of the treatments, only the best treatment will be used, the experimenter will be interested in the direct effect of treatments, and will wish to disentangle this from any carry-over effects. The possibility of carry-over effects is thus a major problem for cross-over designs, and will be considered in detail in chapter 2. At this stage it is sufficient to note that if carry-over

effects are present, the ideal of applying each treatment to an identical experimental unit cannot be achieved, and that, since the potential benefits of cross-over designs depend on the approximation to this ideal, the consequences of carry-over effects can be far-reaching. It is thus important to be able to detect the presence of carry-over effects in a cross-over design, and to have some means of correcting or allowing for the problems that such effects can have on the data analysis. Thus the problem of carry-over will be considered in detail throughout the thesis.

1.5 The Two-perod Cross-over

The simplest cross-over design involves two treatments, A and B, and two treatment periods. In it, subjects are randomly assigned to one of two possible treatment sequences; either treatment A in period one followed by treatment B in period two, or the treatments in the reverse order. This simple two-period cross-over, or 2×2 cross-over, has been an extremely popular design for clinical trials, and is still comonly used, particularly in trials of new therapeutic drugs by the pharmaceutical industry. Hills & Armitage[1979] give a simple explanation of the statistical analysis of this design, but it will be necessary to consider this in greater detail, when it will be seen that carry-over effects are particularly difficult to deal with, because

it is not possible to estimate the size of such effects accurately.

1.6 Multi-period Cross-overs

In most cases, the statistical analysis of a cross-over design will involve positing a linear model for the responses, in which the response is considered to be the sum of effects due to the subjects, periods ,treatments etc. One reason the two-period cross-over has difficulty in coping with carry-over effects is that this model becomes too complex for the relatively small number of observations. An obvious remedy is to increase the number of observations by including extra periods. It is also possible to deal with more than two treatments in a cross-over design. As the number of periods and/or treatments increase, the number of possible treatment sequences increases, and it may become impractical to obtain observations on all possible sequences. The choice of which sequences are to be represented in the design is generally influenced by considerations of the balance that will be achieved, both in terms of which treatments occur in each period, and which treatments precede others. These considerations will be dealt with in more detail in chapter 3.

1.7 Analysis Considerations

As with any experiment, the response variables used to measure the effect of the treatments in a cross-over design may be of various forms. They may be continuous, binary, categorical, or ordered categorical random variables, there may also be a single response variable of interest (univariate) or many (multivariate). Each type of data will, generally speaking, demand a different method of statistical analysis. Still greater diversity of methods is possible depending on the basis of the analysis method. Broadly, the analysis could employ classical parametric methods, non-parametric methods, or Bayesian methods. Most of the combinations of variable type and analysis method have been considered for the two-period cross-over, and Hills & Armitage[1979] refer at least briefly to many. A review of methods for binary data has been given more recently by Kenward & Jones[1987a]. This thesis will consider the analysis of the two-period cross-over with continuous response variables in chapter 4, and the corresponding analyses for selected, more complex designs in chapter 5. The analysis of the two-period cross-over with binary data will be considered in chapter 6, with the corresponding analysis for more complex designs being considered in chapter 7. More recently, a Bayesian analysis for the two-period cross-over has been given by Grieve[1985], and this form of analysis will be considered in chapter 9,

which will also consider the use of Gibbs sampling to perform a Bayesian analysis.

1.8 Discussion

Studies of the literature show that cross-over designs are still widely used, especially for clinical trials. Jones & Kenward[1989] report that a survey of the British Medical Journal between January 1980 and April 1988 found over 80 reports of cross-over trials, most of which used the simple two-period design. They also report that a survey of 12 large pharmaceutical companies in the USA (Fava & Patel[1986]) identified 72 cross-over trials, over half of which used the simple two-period design. This confirms that, in clinical trials at least, the two-period cross-over is still commonly used, in spite of criticisms of its short-comings (see for example Brown[1980]). The aim of this thesis is firstly to consider ways in which the analysis of the two-period cross-over can be improved, for example by including baseline observations, and secondly to examine alternative cross-over designs which are more complicated than the two-period cross-over, but which give more powerful tests and better estimates of the treatment effects.

Chapter 2: Carry-over

2.1 Introduction

In the previous chapter, the advantages of using a subject as his own control in a cross-over design were outlined. It was pointed out, however, that these advantages are the result of assuming that the subject is in the same condition at the start of each treatment period. This will not be the case if treatments which have been applied in previous periods are still having some effect on the subject, so that the subject's initial condition is being modified by these carry-over effects. Thus the presence of carry-over effects can, at least partially, negate the apparent advantages of a cross-over design. It is certainly true that most of the problems associated with cross-over designs arise because of the possibility of carry-over effects, and that there would be few problems if it was known that carry-over effects were impossible. This thesis is concerned with ways of dealing with the problems caused by carry-over effects, so it is necessary to consider the nature of the problems that occur, and the ways in which carry-over effects can arise. The problems caused by carry-over effects will be illustrated by considering the simple two-period cross-over design, and the possible causes of carry-over effects will then be examined.

2.2 The Two-period Cross-over

In the two-period cross-over design, subjects are randomly assigned to one of two treatment groups. Those in the first group receive a period of treatment with treatment A, follwed by an equal period of treatment with treatment B ; while subjects in the second group receive the treatments in the reverse order (B then A). For the purpose of illustrating how the presence of carry-over effects alter the analysis , it will be assumed that a continuous response variable is measured at the conclusion of each treatment period. Then, following Hills & Armitage[1979], a linear model for the response variable can be set up. For the time being, it will be assumed that carry-over effects are not present ,so that the model for y_{ijk}, the observation on subject j in group i for period k, will involve the following terms:

μ = the overall mean

ω_{ij} = the effect of an individual subject

π_k = the effect of a treatment period

τ_L = the effect of a treatment

and ϵ_{ijk} = the random error.

ω and ϵ will be considered as random effects with variances σ_ω^2 and σ_ϵ^2. The ω's and ϵ's are independent, but the two observations on the same subject will be correlated, with $\text{cov}(y_{ij1}, y_{ij2}) = \sigma_\omega^2$. All the other terms are fixed effects, with the usual restrictions that $\pi_1 + \pi_2 = 0$ and $\tau_A + \tau_B = 0$. Because of these restrictions,

reference to the presence of one of these fixed effects in the model should be taken to imply that there is a difference between the two levels of the factor. Applying this model gives the following expressions:

for patient i in sequence 1;

period 1: $y_{1i1} = \mu + \omega_{1i} + \pi_1 + \tau_A + \epsilon_{1i1}$ (2.2.1)

period 2: $y_{1i2} = \mu + \omega_{1i} + \pi_2 + \tau_B + \epsilon_{1i2}$ (2.2.2)

for patient j in sequence 2;

period 1: $y_{2j1} = \mu + \omega_{2j} + \pi_1 + \tau_B + \epsilon_{2j1}$ (2.2.3)

period 2: $y_{2j2} = \mu + \omega_{2j} + \pi_2 + \tau_A + \epsilon_{2j2}$ (2.2.4)

Writing D_{1i} for the difference between the two observations on subject i in group one, $(y_{1i1} - y_{1i2})$ and D_{2j} for the corresponding difference for subject j in group two, it can be seen that $E(D_{1i}) = (\pi_1 - \pi_2) + (\tau_A - \tau_B)$ and $E(D_{2j}) = (\pi_1 - \pi_2) + (\tau_B - \tau_A)$ with $var(D_{1i}) = var(D_{2j}) = 2\sigma_e^2$

D_{1i}, D_{2j} are within subject comparisons, being the comparison of the two observations on the same subject, and as such are less variable than a between subject comparison. Since observations on different subjects are uncorrelated, a similar between-subject comparison (e.g. $y_{1i1} - y_{1j2}$) would have a variance of $2(\sigma_s^2 + \sigma_e^2)$. As the between subject variation σ_s^2 is usually larger than the within subject variation σ_e^2, it can be expected that the

the between-subject comparison will have a variance that is more than twice that of the within-subject comparison.

Clearly, if $\tau_A = \tau_B$, the expected values of D_{1i} and D_{2j} are equal, so that a two-sample test comparing the D_{1i} and the D_{2j} gives a test of the hypothesis $\tau_A = \tau_B$. If it is assumed that there are n_i subjects in group i and that $\bar{D}_{i.} = (\Sigma D_{1i})/n_1$ and $\bar{D}_{2.} = (\Sigma D_{2j})/n_2$; $\frac{1}{2}(\bar{D}_{1.} - \bar{D}_{2.})$ gives an estimate of $\tau_A - \tau_B$ which has a variance of $\frac{1}{2}\sigma^2\{(1/n_1) + (1/n_2)\}$.

2.3 Effect of Carry-over on the Model

Carry-over effects are now introduced into the model with the term α_L denoting the effect of treatment L in the period immediately following that in which it was applied. α_A, α_B are fixed effects, so the restriction $\alpha_A + \alpha_B = 0$ may be used. Once again, this restriction implies that, if carry-over effects are present, the carry-over effects of the two treatments are different.

There will be no carry-over effects in the first period, so the expressions given above will still apply, but the expressions for the observations in the second period will now be as follows:

for patient i in sequence 1;

period 2: $y_{1i2} = \mu + \omega_{1i} + \pi_2 + \tau_B + \alpha_A + \epsilon_{1i2}$ (2.3.1)

for patient j in sequence 2;

period 2: $y_{2j2} = \mu + \omega_{2j} + \pi_2 + \tau_A + \alpha_B + \epsilon_{2j2}$ (2.3.2)

With D_{1i}, D_{2j} defined as before, the expectations are now $E(D_{1i}) = (\pi_1 - \pi_2) + (\tau_A - \tau_B) - \alpha_A$ and $E(D_{2j}) = (\pi_1 - \pi_2) + (\tau_B - \tau_A) - \alpha_B$, so that, even if $\tau_A = \tau_B$, the expectations are not equal and $\frac{1}{2}(\bar{D}_{1.} - \bar{D}_{2.})$ estimates $(\tau_A - \tau_B) - \frac{1}{2}(\alpha_A - \alpha_B)$.

Grizzle[1965] showed by consideration of the least squares equations that the only way of testing the hypothesis $\tau_A = \tau_B$, or obtaining an unbiased estimate of $\tau_A - \tau_B$ is by using the observations from the first period only, which are unaffected by carry-over. It can be seen that $E(y_{1i1}) = \mu + \pi_1 + \tau_A$ and $E(y_{2j1}) = \mu + \pi_1 + \tau_B$ so that the two sets of observations have the same mean if $\tau_A = \tau_B$, and a two-sample test comparing the y_{1i1} and the y_{2j1} will test the hypothesis $\tau_A = \tau_B$. Also, if $\bar{y}_{1.1} = (\Sigma y_{1i1})/n_1$, and $\bar{y}_{2.1} = (\Sigma y_{2j1})/n_2$, $\bar{y}_{1.1} - \bar{y}_{2.1}$ gives an unbiased estimate of $\tau_A - \tau_B$ with variance $(\sigma_s^2 + \sigma_e^2)\{(1/n_1) + (1/n_2)\}$. If the between-subject variation σ_s^2 is larger than the within-subject variation σ_e^2, this variance will be more than four times as large as the variance of the corresponding estimate obtained by using the within-subject comparison when no carry-over was present. Hence the introduction of carry-over effects into the model prevents the use of each subject as his own control, and forces the experimenter to use only the first period observations, reducing the cross-over to a parallel design. This drawback of the two-period cross-over has been known for many years, and in 1977 the

Biometrics and Epidemiology Methodology Advisory Committee (BEMAC) of the U.S. Food and Drug Administration produced a report criticising the use of the two-period cross-over in clinical trials on the grounds that the usual within-subject estimate of the difference between treatment effects is biased if carry-over effects are present. Since there is no benefit in using the design if carry-over effects are present, the FDA advised that it should only be used if there was good reason to believe that carry-over effects would not be present. Despite this discouragement, the design has still been commonly used, at least in the U.K.. It may be that experimenters have disregarded the problem because they believe that carry-over effects can be avoided by judicious organisation of the experiment, or it may be that they have considered that, even if carry-over effects are present, useful information can be obtained from the first period observations. The belief that carry-over can be avoided depends partly on a misconception of the nature of carry-over effects, which will be discussed later in this chapter, while the ability to obtain an unbiased estimate of the difference between treatment effects depends on being able to detect the presence of carry-over.

2.4 Testing for Carry-over

Using the model with carry-over as above, the sum of the observations on each subject is calculated. For subject i in group one, $S_{1i} = y_{1i1} + y_{1i2}$, and for subject j in group two, $S_{2j} = y_{2j1} + y_{2j2}$. Using the restrictions $\pi_1 + \pi_2 = \tau_A + \tau_B = 0$, these have expectations $E(S_{1i}) = 2\mu + \alpha_A$, $E(S_{2j}) = 2\mu + \alpha_B$, and variances $4\sigma_a^2 + 2\sigma_e^2$. Clearly, S_{1i} and S_{2j} have the same expectation if $\alpha_A = \alpha_B$, so that a two sample test comparing the S_{1i} and the S_{2j} gives a test of $\alpha_A = \alpha_B$. However, this is clearly a between-subject comparison, and due to the large variances of S_{1i} and S_{2j}, the test will not be very powerful. The lack of power is emphasised by noting that the corresponding estimate of $\alpha_A - \alpha_B$, obtained from $\bar{S}_{1.} - \bar{S}_{2.}$ where $\bar{S}_{1.} = (\Sigma S_{1i})/n_1$ and $\bar{S}_{2.} = (\Sigma S_{2j})/n_2$, has a variance of $(4\sigma_a^2 + 2\sigma_e^2)\{(1/n_1) + (1/n_2)\}$. If $\sigma_a^2 > \sigma_e^2$, this variance will be more than three times that of the estimate of treatment differences from the first period only, and more than twelve times that of the estimate of treatment differences obtained from both treatment periods. The result of this is that small differences in carry-over effects may go undetected, and lead to the use of a test which is strictly speaking invalid, and the use of estimates for the difference between treatments which are biased, being contaminated with the undetected carry-over effects.

It should also be noted that by comparing the sums of the observations on subjects in groups one and two the groups are in effect being compared. The linear model given above did not contain terms for the effect of the groups, although these are implicitly contained in the set of individual subject effects. The sum of squares for differences between subjects could be split into a sum of squares for differences between the groups, and a sum of squares for subjects within groups. Should there be a systematic difference between the two groups, this will be confounded with the carry-over effects. It is usually argued that, because subjects are allocated at random to the two groups, there should not be any systematic difference between them. However, it should not be forgotten that the groups are treated differently, by being given different sequences of treatments, and this might induce a difference between them. Any difference between the sequences would of course be confounded with differences between the groups, and would be detected by the test for carry-over given above. Thus the standard test for carry-over in the two-period cross-over is not only a weak test, but also a test which is not specifically aimed at the carry-over effect.

2.5 The Nature of Carry-over Effects

The name "carry-over" suggests an effect of a treatment which is left over from the period in which it

was applied. The most obvious such effect would occur if a drug which had been applied in a previous period was still present in the body, and having some effect, in a subsequent period. Such carry-over is usually called "pharmacological", and must be very rare. Experimenters will as a rule be knowledgeable about the bioavailability of any drug they use, and ensure that such pharmacological carry-over will not occur. In some cross-over trials, a wash-out period, in which no treatment is applied, is inserted between the treatment periods so that all trace of the drug administered previously will disappear. Experimenters can thus be confident that pharmacological carry-over can be avoided, and it may be that some have been too ready to extend this confidence to all forms of carry-over.

If carry-over is not pharmacological in nature, it may be psychological. A particularly good, or bad, first treatment could change the subject's expectations about the second treatment, and subtly change their reaction to it. This is most likely to affect the results if the response variable is an assessment by the patient of how he feels, but given the effects that can arise from placebo treatments, it would be unwise to dismiss the possibility of such psychological effects on objectively measured response variables.

Although pharmacological and psychological carry-over effects are the most straight-forward, they are by

no means the only possibilities for effects that can disrupt the simple analysis of a cross-over design. Any effect which alters the response of a given treatment administered at different times will do this. Considered thus, it is natural to regard such effects as treatment-period interactions, and this has become the preferred description for researchers who wish to avoid the pharmacological and psychological implications of the term "carry-over". The more general nature of the term "treatment-period interaction" does indeed convey the sense that any effect causing differences in response to the same treatment at different times is to be included, and is more generally applicable to cross-over designs with more than two periods or more than two treatments. In the simple two-period cross-over, with two treatments and two periods, there will be one degree of freedom for treatment-period interaction, as there is for carry-over. Since these two effects are alternative ways of explaining how a treatment can have different effects in different periods, it is not surprising that they are completely equivalent to one another. For a more complex cross-over design with, say, three treatments and three periods, there would be four degrees of freedom for treatment-period interaction. Each of the three treatments could have a "first-order" carry-over effect, which would affect the treatment immediately following, but these would only take up two degrees of freedom. The

other two degrees of freedom could then be accommodated by allowing the three treatments to have second-order carry-over effects, which would affect the treatment applied in two periods time. Thus first and second order carry-over effects together would be equivalent to treatment-period interaction. By including higher-order carry-over effects in the model in this way, an equivalent to treatment-period interaction can be constructed for any cross-over design

While the name carry-over has led people to think in terms of easily avoided pharmacological effects, the term treatment-period interaction has also induced a particular mode of thinking, leading some authors (e.g. Kenward & Jones[1987a]) to worry about breaching marginality rules (Nelder[1977, 1982]), by fitting the interaction without both of the main effects. If the effect is in fact a pharmacological or psychological effect which it is reasonable to regard as an effect of the previous treatment, there is no reason to consider it as an interaction and be worried about marginality.

In addition, although treatment-period interaction seems to be a more general term, there are circumstances in which it does not seem appropriate. Consider, for instance, an experiment in which a wash-out period is inserted between the two treatment periods, and an observation is taken at the end of the wash-out period. If pharmacological or psychological carry-over is

present, this observation will be modified by that effect, so that it will be different from a baseline observation taken at the end of a run-in period. However, it does not seem sensible to talk of a treatment-period interaction in this case, because no treatment is applied in the wash-out period.

Considerations such as this have led some authors (e.g. Huitson et al[1982]) to talk of both carry-over and treatment-period interaction, presumably meaning by carry-over those effects which can be thought of as being caused by the previous treatment, and by treatment-period interaction, any effects which cannot be directly ascribed to the treatment. Although it seems reasonable to draw a distinction in this way, the two sorts of effect are confounded. If the effect of a treatment is found to differ in the different periods of a cross-over design there will be no way of knowing whether this is due to "carry-over" from a previous treatment, or less specific treatment-period interaction. As it is impossible to differentiate between the two types of effect, and both have the same undesirable effect on the analysis, it is more usual to use only one of the two possible names for these effects, giving rise to the possibility of the misconceptions that seem to be inherent in the two appelations.

Chapter 3: Higher-order Cross-overs

3.1 Introduction

In the last chapter, the consequences of carry-over effects on the two-period cross-over were examined. These were severe, and it is clear that, with carry-over effects in the model, the complexities are too great for the relatively small number of observations. An obvious way of attempting to cope with this problem is to increase the number of observations. The simplest way of doing this is to add baseline readings, which can be quite effective, as will be seen in chapter four. In this chapter the more radical alternatives of increasing the number of periods, sequences or treatments, or some combination of these, will also be considered.

3.2 Increasing the Number of Sequences

Given two treatments and two periods, there are two choices of treatment for each period and hence four possible sequences; AA, BB, AB and BA. It could be argued that only the sequences AB and BA are eligible for inclusion in a cross-over trial, because there is no change over from one treatment to another in the other two sequences. However, this seems unnecessarily restrictive. It would be quite possible to arrange a trial with sequences AA and BB to mimic a cross-over trial, and, if the usual conditions for blindness were

met, the subject and physician would believe that the treatment in the first and second periods were different. It would then seem reasonable to regard such a trial as a cross-over trial, and to analyse it as though it were one, by using a linear model similar to that given in section 2.2.

Clearly, a trial involving only the two sequences AA and BB would effectively be a parallel trial, and so is of little interest. However, a trial involving all four possible sequences, AA, BB, AB and BA would have useful properties. The observations taken at the end of the two treatment periods for subjects receiving the sequence AA or BB would be expected to be the same, except for differences in period effects, and any carry-over effect. Thus these sequences will give information about the carry-over effects which will allow a better test of carry-over, and a within-subject comparison for treatment effects to be made. This assumes that the carry-over effect of a particular treatment is always the same, and does not depend on the following treatment in any way. The analysis of this design is considered in detail in chapter 4. It has been shown by Laska et al.[1983] that this design, with equal numbers in the four sequences, is the optimal two-period design, giving unbiased estimators of the treatment difference which have minimum variance, when carry-over is present.

3.3 Increasing the Number of Treatment Periods

3.3.1 Introduction

Even if a clinical trial is comparing two treatments, there is no reason why a design with more than two treatment periods should not be used. Clearly, as the number of treatment periods increases the number of possible sequences increases, with 2^p sequences for p periods. Although an experiment could be carried out in which all possible sequences are represented, it is likely that it would be unnecessary or inefficient to do this. There is a long history of literature which considers the desirable properties of cross-over designs, often considering designs with more than two treatments, and methods of constructing designs with these desirable properties. As mentioned in chapter 1, it is frequently the case that the response variable will change with time, as with the variation in milk yield over a cow's lactation cycle, or the diminution of post-operative pain with the passage of time, so that observations taken in different periods could be expected to be different. Experimenters realised from the outset that the presence of such period effects necessitated the use of a balanced design, in which each treatment occurred equally often in each of the treatment periods. This could easily be achieved by using a latin square design with rows as treatment periods and columns as experimental units.

Where two treatments are involved, this balance can be achieved by using sequences which form complementary pairs. In such a complementary pair, one sequence can be obtained from the other by interchanging the two treatments (e.g. ABAAB and BABBA). Thus the simple two-period cross-over consists of a single complementary pair of sequences (AB and BA). If equal numbers of subjects are allocated to each of the two complementary sequences, the design is said to be "dual balanced" (Matthews[1987]).

In general, experimenters will be reluctant to consider a large increase in the number of periods, as this would increase the length and cost of the experiment. Kershner & Federer[1981] have examined most of the possible designs with two, three and four treatment periods, comparing the variances of the estimates of the difference between treatments and first-order carry-over effects. Table 1 contains an extract from Kershner & Federer's results, giving the leading factors in the variances of the unbiased estimates of contrasts between the treatment and carry-over effects for various two-treatment cross-over designs. Because Kershner & Federer use a model containing sequence, or group effects, their table indicates that estimates of the treatment and carry-over differences are not possible with the simple two-period cross-over. The table shows that the three-period design with sequences ABB and BAA is particularly efficient in giving estimates with a relatively low

variance. This design is of particular interest, and will be dealt with extensively in this thesis. At this stage it is interesting to consider the properties that make it efficient.

Table 1 Leading factors of the variances of unbiased estimators of contrasts between treatment and carry-over effects for various two-treatment cross-over designs.

Design Sequences	Leading factor of variance	
	Treatments	Carry-over
AB, BA	NE	NE
AB, BA, AA, BB	8.00	16.00
AAB, BBA	2.00	8.00
ABB, BAA	1.50	2.00
ABA, BAB	6.00	8.00
AAB, BBA, ABB, BAA	1.55	3.10
AAB, BBA, ABA, BAB	3.00	6.00
ABB, BBA, ABA, BAB	1.85	2.46
AAB, BBA, ABB, BBA, ABA, BAB	1.85	3.18
ABAB, BABA	5.50	8.00
ABBB, BAAA	1.38	1.50
AABB, BBAA, ABBA, BAAB	1.00	1.45

(NE = not estimable)

3.3.2 Efficiency of three -period Designs

The eight possible sequences of two treatments in three periods form four complementary pairs : AAA, BBB; AAB, BBA; ABA, BAB and ABB, BAA. The first of these pairs is effectively a parallel design, and will not be considered, leaving three possible designs, each consisting of one complementary pair. Clearly, each of these is unbalanced in so far as each subject does not receive an equal amount of time on the two treatments, but there is nothing to choose between the three in this respect. It is thus necessary to consider how first-order carry-over affects the three possible designs. In the first pair (AAB, BBA) subjects in sequence one only experience carry-over from treatment A while those in sequence two only experience carry-over from B. Thus carry-over will be confounded with sequences, and, because one treatment is more associated with each sequence, it could be expected that the tests for carry-over and treatments would be correlated. This can in fact be shown to be the case.

In the second pair (ABA, BAB) first-order carry-over from A & B occurs in each sequence, but carry-over from A always acts on B and vice versa, suggesting that, for this design also, tests for treatment and carry-over will be correlated. This can also be shown to be true. The third pair (ABB, BAA) has neither of these problems, both sequences experiencing carry-over from each treatment,

and each carry-over acting on each of the treatments. This is the design that is particularly efficient, and it can be shown that the tests for treatments and carry-over are independent. The analysis of this design will be considered in detail in chapters 5,7 and 9.

3.3.3 Properties of Efficient Designs

Although it is relatively easy to see that the ABB; BAA design is free from the deficiencies of the alternatives, it is less easy to see what properties it has that are characteristic of efficient designs. In one of the earliest papers on cross-over designs, Cochran et al.[1941] not only realised that period effects meant that treatments should appear equally often in each period, but also that the possibility of carry-over effects imposed the need for another sort of balance, in which each treatment was preceeded by each other treatment equally often. The design they employed consisted of two 3x3 latin squares giving six sequences: ABC, BCA, CAB, ACB, BAC, CBA. It can be seen that, in this design, each treatment occurs twice in each of the three periods, and is immediately preceeded twice by each other treatment. The design allows a correction for possible carry-over effects to be made, but the authors point out that with two treatments, "to obtain direct information on carry-over effects ... it would be necessary to include units receiving the sequences AA &

BB, as well as units receiving the switch-over sequences AB & BA".

Later authors generally took for granted that the two sorts of balance mentioned by Cochran et al. were desirable, and sought to give rules by which designs with such balance could be obtained, although some extended or specialised the balance requirements. Thus Williams[1949,1950] considered designs balanced for pairs of residual (i.e. carry-over) effects, while Berenblut[1964,1967,1968] considers designs which are balanced for the linear component of the carry-over effects, for a quantitative factor at equally-spaced levels, and more recently, designs for use with autocorrelated errors have been considered (Bora[1984], Matthews[1987]).

A typical set of conditions for the construction of cross-over designs is due to Patterson[1952]. Patterson was considering designs with more than two treatments, but it is nevertheless interesting to consider his conditions in respect to the three designs outlined above. Patterson gave seven conditions, the first three of which he considered necessary if carry-over effects were not present, with all seven to be used if carry-over was present. His conditions are as follows:

I. No treatment occurs in a given sequence more than once.

II. Each treatment occurs in a given period an equal number of times.

III. Every two treatments occur together in the same number of sequences.

IV. Each ordered succession of two treatments occurs equally often.

V. Every two treatments occur together in the same number of curtailed sequences, formed by omitting the final period.

VI. In those sequences in which a given treatment occurs in the final period the other treatments occur equally often.

VII. In those sequences in which a given treatment occurs in any but the final period, each other treatment occurs equally often in the final period.

Condition I is clearly not met by any of the three designs being considered, while conditions II and III are met by them all, and would be met by any design using sets of complementary sequences. Condition IV also appears to be met by all three designs, as, given condition I, Patterson clearly meant by "ordered succession of two treatments" a succession of two different treatments. This reflects the earlier requirement of Cochrane et al. for each treatment to be preceeded equally often by each other treatment. As noted earlier, Cochrane et al. also realised the need for sequences AA & BB if carry-over effects are to be

estimated, but in spite of this, it was not until the paper on extra-period latin sqare designs by Lucas[1957] that the advantages of a treatment preceeding itself as often as other treatments preceeded it was generally appreciated.

In an extra-period latin square design an extra period is added, and the final treatment repeated. With a 2x2 latin square this results in the design with sequences ABB & BAA. Lucas points out that first-order carry-over effects and treatment effects are then orthogonal. This idea was later combined with Patterson's rules in a joint paper (Patterson & Lucas[1959]), but it is only necessary to interpret Patterson's Condition I more liberally, to include any possible ordered pair of treatments i.e. AA, BB, AB or BA, to obtain the benefit of this insight. The design with sequences ABA & BAB does not satisfy the condition as the successions AA & BB do not occur. The other two designs do, however, satisfy the modified condition. Of these two designs, that with sequences AAB & BBA does not satisfy any of Conditions V, VI and VII, while the efficient design, with sequences ABB & BAA, satisfies all three. The design with sequences ABA & BAB, which does not satisfy the modified Condition IV also satisfies these last three conditions. Thus it seems that, for a design consisting of sets of complementary sequences, a good design will satisfy the modified Condition IV and Conditions V, VI and VII. It is

interesting to note that the design with the four sequences AA, BB, AB and BA also satisfies these four conditions, except that truncated sequences only consist of one period so that condition V is redundant.

3.3.4 Optimal Designs

In determining the optimum cross-over designs, giving the best (minimum variance) linear unbiased estimators of the difference between treatments and first-order carry-over effects, Laska et al.[1983], Laska & Meisner[1985] also give conditions that will be met by optimum designs. These conditions relate to ideas of "uniformity" and "balance" that the authors define. A design is said to be "uniform on the periods" if, in each period, the same number of patients are assigned to each treatment; and "uniform on the patients" if, for each patient, each treatment appears in the same number of periods. A "uniform" design is one which is uniform on the periods and on the patients. A design is defined as "balanced" if each treatment is preceded by each of the other treatments equally often, and "strongly balanced" if each treatment is immediately preceded by each of the treatments, including itself, equally often. It is then stated that a strongly balanced uniform design will be optimal whether or not first-order carry-over effects are present, and whether or not baseline observations are taken. Clearly designs with an odd number of treatment

periods cannot be uniform on patients, so that no such strongly balanced uniform design exists. However, the authors also state that, in these circumstances, if the truncated design is a strongly balanced uniform design, and the final treatment in each sequence is the same as the penultimate treatment, the design will be optimal, again, whether or not first-order carry-over effects are present and whether or not baseline observations are taken. These rules can be used to show that for two periods and two treatments the design with the four sequences AA, BB, AB and BA is optimal, and that with three periods and two treatments, the design with the two sequences ABB and BAA is optimal.

It should be noted that designs consisting of sets of complementary pairs of sequences will be uniform on periods if the same number of subjects are allocated to each sequence in a complementary pair, while the concept of strong balance is effectively equivalent to Patterson's Condition IV, as modified above. Patterson's Conditions V, VI and VII are necessary to cope with designs with an odd number of treatment periods, and are approximately equivalent to Laska's second rule. It should also be remembered that, in a design with more than two treatment periods, first-order carry-over alone is not equivalent to treatment-period interaction. If treatment-period interaction, or the equivalent set of higher-order carry-over effects, was included, the rules

for optimality might need to be modified. Lucas[1957] suggested that "more than one extra period may be called for when residual (i.e. carry-over) effects are assumed to continue for more than one period"

3.4 Designs to Consider Further

In the remaining chapters of this thesis four designs have been singled out for detailed examination. The first of these is the simple two-period cross-over, which is the design that has been most commonly used in clinical trials, and is included as a comparison for the other designs. The second design is the one identified by Laska et al. as the optimum two-period design, with four sequences AA, BB, AB and BA; while the third is their optimum three-period design, with two sequences ABB and BAA. As has already been mentioned, only first-order carry-over effects were considered in identifying this as the optimum three-period design, and, by analogy with the two-period case, it might be expected that the design with four sequences AAA, BBB, ABB and BAA, might prove best if higher-order carry-over is included to obtain equivalence with treatment-period interaction, and this three-period, four sequence design is included in the detailed study. It is also clear that this design, and the two-period design with four sequences, are effectively a combination of a parallel and a cross-over design. It will thus be useful to investigate whether the

combination is an improvement over the parallel or cross-over designs separately.

Chapter 4: Classical Analysis for Continuous Response Variables: Two-period Designs

4.1 Introduction

In the previous chapter, four cross-over designs were identified as worthy of further study. These are the simple two-period cross-over, included mainly for comparison with the other designs, the "complete" two-period cross-over, consisting of all four possible sequences: AA, BB, AB & BA; the three-period design with sequences ABB & BAA, identified by Kershner & Federer as being efficient; and the three period design with sequences AAA, BBB, ABB & BAA, which combines the features of this design and the parallel design. In this chapter and the next, the analysis of these designs will be considered assuming there is a single response variable which is a continuous random variable. The analysis of the above two-period designs will be considered in the present chapter, while the three-period designs will be dealt with in chapter 5. Classical, as opposed to Bayesian methods are considered in these chapters. The tests available in the classical framework could be parametric, mainly t-tests, or non-parametric equivalents to t-tests such as the Mann-Whitney U-test. Bayesian methods, and methods for response variables which are not continuous, will be considered in later chapters.

The linear model which will form the basis of the analysis in this chapter was introduced in chapter 2, and the simple analysis of the two-period cross-over has also been given in that chapter. In the present chapter this analysis will be extended to cover the situation where baseline observations are available, and corresponding analyses will be given for the three other designs.

4.2 The Simple Two-period Cross-over

Analysis of the two-period cross-over was considered in section 2.1, and will not be repeated here. However, it is necessary to consider how the analysis can be improved by the inclusion of baseline measurements. The term "baseline" is used here to mean an observation taken at the start of a clinical trial, before treatment begins. It should be noted that some authors use the term in a rather different way. When Kershner & Federer[1981] compare two-treatment cross-over designs, they consider the designs with and without baselines, but take "with baselines" to mean that each observation for a treatment period is preceded by a "baseline" observation. This is also the case in the paper by Kenward & Jones[1987b] covering the analysis of the simple two-period cross-over when "baseline" measurements are available. In this thesis, only the observation preceding the first treatment period will be called a "baseline", and an

observation between the first and second treatment periods will be called a "washout" observation.

Using the linear model outlined in chapter 2, expressions can be written down for the baseline observations, y_{ij0}. As no treatment is applied in the baseline periods, these expressions will not involve a treatment effect, but only a period effect, labelled π_0, a subject effect ω_{ij}, and an error term ϵ_{ij0}. The extra information in the baseline observations allows the model to be complicated by the introduction of group or sequence effects γ_1 and γ_2. The individual subject or patient effects ω_{ij} must be regarded as deviations from the sequence effects, so that the sum of the patient effects within each group is zero. The sequence effects will be regarded as fixed effects, with the restriction $\gamma_1 + \gamma_2 = 0$, while the patient effects will again be regarded as random effects. Using τ_A, τ_B and α_A, α_B to represent the treatment and carry-over effects in the model, the expressions for the observations are:

for patient i in sequence 1;

baseline: $y_{i10} = \mu + \gamma_1 + \omega_{i1} + \pi_0 + \epsilon_{i10}$

period 1: $y_{i11} = \mu + \gamma_1 + \omega_{i1} + \pi_1 + \tau_A + \epsilon_{i11}$

period 2: $y_{i12} = \mu + \gamma_1 + \omega_{i1} + \pi_2 + \tau_B + \alpha_A + \epsilon_{i12}$

for patient j in sequence 2;

baseline: $y_{2j0} = \mu + \gamma_2 + \omega_{2j} + \pi_0 + \epsilon_{2j0}$

period 1: $y_{2j1} = \mu + \gamma_2 + \omega_{2j} + \pi_1 + \tau_B + \epsilon_{2j1}$

period 2: $y_{2j2} = \mu + \gamma_2 + \omega_{2j} + \pi_2 + \tau_A + \alpha_B + \epsilon_{2j2}$

..4.2.1

The expected values of the baseline observations for the two sequences differ only in the sequence or group parameters, so that using a two sample test to compare the two sets of baseline observations will test whether there is any systematic difference between the two groups, or whether the random allocation of subjects to the two groups has been successful in avoiding such a difference. This test would probably be the first carried out on a data set of this type, since one would normally wish there to be no significant difference between the groups. The sequence of tests would probably continue by testing for carry-over, as the test for treatments will depend on whether there is significant carry-over. The presence of baseline measurements allows a better test of carry-over. The combination $E_{1j} = 2y_{1j0} - y_{1j1} - y_{1j2}$ has expected value $(2\pi_0 - \pi_1 - \pi_2) - \alpha_A$ for sequence 1 and $(2\pi_0 - \pi_1 - \pi_2) - \alpha_B$ for sequence 2. Thus a comparison of the two sets of combinations E_{1i} and E_{2j} using a two sample test will test whether there is any difference in the two carry-over effects. The E_{ij}'s are of course within-subject comparisons, from which the individual

subject effects have been eliminated, so that this test of carry-over will be more efficient than the one given in section 2.1. This can be seen by considering the variance of the estimate of the difference in carry-over effects $\alpha_A - \alpha_B$ which can be obtained from the E_{ij}'s. Writing $\bar{E}_{1.}$ for the average of the n_1 E_{1i} values and $\bar{E}_{2.}$ for the average of the n_2 E_{2j} values, $\bar{E}_{2.} - \bar{E}_{1.}$ is an unbiased estimator of $\alpha_A - \alpha_B$ with variance $6\sigma_e^2(1/n_1 + 1/n_2)$. It can be seen from table 2 below that this is only smaller than the corresponding variance for the estimator given in chapter 2 if the within-subject variation σ_e^2 is smaller than the between-subject variation, σ_s^2, but this will normally be the case.

If there is a significant difference in the carry-over effects, the data from the second period must be discarded, but it is still possible to obtain a within-subject comparison between the baseline and first-period observations which can be used to test for a difference between the treatment effects. Writing $F_{1i} = y_{1i0} - y_{1i1}$ and $F_{2j} = y_{2j0} - y_{2j1}$, it can be seen that the expected value of F_{1i} is $(\pi_0 - \pi_1) - \tau_A$, while the expected value of F_{2j} is $(\pi_0 - \pi_1) - \tau_B$, so that a two sample test comparing the two sets of F's will test for difference between the treatment effects, while the corresponding estimate of $\tau_A - \tau_B$ is $\bar{F}_{2.} - \bar{F}_{1.}$, where $\bar{F}_{i.}$ is the mean of the n_i F_{ij} values. This estimate has a variance of $2\sigma_e^2(1/n_1 + 1/n_2)$.

If the test for carry-over is non-significant, so that the data from the second treatment period need not be discarded, the usual test of treatment differences using the contrast of the two treatment periods in the two sequences given in 2.1 can be used. It can be seen from table 2 that this test yields an estimate of the treatment difference that has a variance of one quarter of that given by the comparison of the baseline and first-period observations. It is thus still very beneficial for there to be no carry-over in the experiment, although since the test of carry-over is more powerful with baseline data, there will be less likelihood of making an error which leads to the use of the inappropriate test for treatments. The position is still complex, however, as will be seen when the question of power is examined in chapter 8.

If there is no carry-over, the estimate of the treatment difference based on the F's is of course still valid, and it might seem possible to obtain a still better estimator by combining the estimates based on the D's and the F's. This is not possible, however, because the two estimators are positively correlated, and the variance of a combined estimator will be larger than that based on the D's only. On the other hand, if carry-over is present, the estimator based on the D's will be biased, having an expected value of $\tau_A - \tau_B - \frac{1}{2}(\alpha_A - \alpha_B)$. Since $\alpha_A - \alpha_B$ can be estimated from the E's, it might

seem possible to correct for the bias in the D-estimator, and obtain an unbiased estimator. This would be $\frac{1}{2}(\bar{D}_{1.} - \bar{D}_{2.}) + \frac{1}{2}(\bar{E}_{2.} - \bar{E}_{1.})$. However, this merely reduces to the F-estimator $\bar{F}_{2.} - \bar{F}_{1.}$.

Table 2 variances of estimators for the two-period cross-over design.

Estimators of $\tau_A - \tau_B$

Combination	Estimator	Variance[1]
$D_{ij} = y_{ij1} - y_{ij2}$	$\frac{1}{2}(\bar{D}_{1.} - \bar{D}_{2.})$	$\frac{1}{2}\sigma_d^2$ [2]
	$\bar{y}_{1.1} - \bar{y}_{2.1}$	$\sigma_w^2 + \sigma_d^2$
$F_{ij} = y_{ij0} - y_{ij1}$	$\bar{F}_{2.} - \bar{F}_{1.}$	$2\sigma_d^2$

Estimators of $\alpha_A - \alpha_B$

Combination	Estimator	Variance[1]
$S_{ij} = y_{ij1} + y_{ij2}$	$\bar{S}_{1.} - \bar{S}_{2.}$	$4\sigma_w^2 + 2\sigma_d^2$
$E_{ij} = 2y_{ij0} - y_{ij1} - y_{ij2}$	$\bar{E}_{2.} - \bar{E}_{1.}$	$6\sigma_d^2$

Notes: [1] all variances are multiplied by $(1/n_1 + 1/n_2)$

[2] this estimator is only unbiased if there is no difference in carry-over effects

Although adding baseline observations is a very simple modification of the usual analysis of the two-

period cross-over design, it nevertheless much improves the efficiency of the two period cross-over by allowing a within-subject comparison to be used for the test of carry-over, and for treatments if carry-over is significant. Table 2 gives the variances of estimators which may be used with the two-period cross-over, from which the improvement gained by the use of baselines can be seen.

Sometimes an experiment is organised with a washout period in between the first and second treatment periods in order to avoid pharmacological carry-over. The opportunity then exists for a second "baseline" observation to be taken at the end of the washout period, before the second treatment period begins. To avoid confusion this will be referred to as the washout observation. It is clearly possible to write down an expression in the form of a linear model for this washout observation, although there is some debate about how the effect of carry-over should be parameterised. It seems clear that the treatment applied in the first period could have some effect on the response recorded as the washout observation, so that the model for the washout observation should include a carry-over term. It seems equally uncontroversial to assume that there will be no carry-over from the wash-out period itself to the second treatment period, as no treatment is being administered in the washout period. However, it may be that the

treatment given in the first period could still have an effect on the second treatment period, although this may well not be the same effect as was observed in the washout period, or would have been observed if the second treatment period had immediately followed the first. In order to allow for this the terms α_A and α_B will be used for carry-over effects from one period to the next i.e. from the first treatment period to the washout period, and β_A, β_B for carry-over from the first period to the second treatment period, when these are separated be a washout period. In order to retain the second treatment period as period two, the wash-out period between the two treatment periods will be labelled period 1'. The expressions for the washout and second treatment period observations are then:

for subject i in sequence 1:

washout : $y_{i1'} = \mu + \gamma_1 + \omega_{i1} + \pi_{1'} + \alpha_A + \epsilon_{i1'}$

period 2: $y_{i2} = \mu + \gamma_1 + \omega_{i1} + \pi_2 + \tau_B + \beta_A + \epsilon_{i2}$

for subject j in sequence 2:

washout : $y_{j1'} = \mu + \gamma_2 + \omega_{j2} + \pi_{1'} + \alpha_B + \epsilon_{j1'}$

period 2: $y_{j2} = \mu + \gamma_2 + \omega_{j2} + \pi_2 + \tau_A + \beta_B + \epsilon_{j2}$ .. 4.2.2

Tests for differences in both types of carry-over in the model are possible using within-subject comparisons. The test of $\alpha_A = \alpha_B = 0$ can be made by means of the

within-subject comparison $F'_{ij} = y_{ij0} - y_{ij1}$. The expected value of this comparison for the two sequences is:

$$E\{F'_{1j}\} = (\pi_0 - \pi_{1\cdot}) - \alpha_A$$

$$E\{F'_{2j}\} = (\pi_0 - \pi_{1\cdot}) - \alpha_B$$

Hence, a two sample test comparing the two sets of F' values is a test of $\alpha_A = \alpha_B$. Similarly, a test of the hypothesis $\beta_A = \beta_B = 0$ can be made by using the within-subject comparison $E_{ij} = 2y_{ij0} - y_{ij1} - y_{ij2}$. This has the following expected values for the two sequences:

$$E\{E_{1j}\} = (2\pi_0 - \pi_1 - \pi_2) - \beta_A$$

$$E\{E_{2j}\} = (2\pi_0 - \pi_1 - \pi_2) - \beta_B$$

So that a two sample test comparing the two sets of E's will test the hypothesis $\beta_A = \beta_B$. It is only if this second test is significant that the data from the second treatment period must be discarded, since only the β terms feature in the model for the second treatment period. However, many experimenters would be inclined to regard the carry-over to the second treatment period as a diminished version of the carry-over into the washout period, i.e. to regard the β terms as some unknown fraction of the α terms. Given this view, an experimenter would tend to be somewhat unhappy about the data from the second treatment period if the test for difference in the

α's was significant even if the test for the β's was not, taking this as an indication that carry-over is present in the experiment, but cannot be detected in the second treatment period data. From table 2, it can be seen that the estimate of $\alpha_A - \alpha_B$, obtained from the E_{ij}'s, is much more precise than the estimate of $\beta_A - \beta_B$, obtained from the F_{ij}'s; the variance of the former estimator being one third of the latter. This difference in precision will also be reflected in the power of the corresponding tests, so that a difference in first-order carry-over is much more likely to be detected than a difference in second-order carry-over.

4.3 The "Complete" Two-period Cross-over

The design with four sequences AA, BB, AB, BA is also one of those considered by Kershner & Federer[1981] while more recently, Hüsler & Lienert[1985] have considered its merits. The design is, in fact, a special case of a class of designs considered by Balaam[1968]. Hüsler & Lienert refer to the design as an "incremental" cross-over design to emphasise that the analysis of the design involves examining the difference between the first and second period observations for the four sequences. The parameters used in the linear model for the simple two-period cross-over can be employed again to write down expressions for the observations in this design. The only change will be that there are now four

sequences and so four sequence effects γ_1, .., γ_4. It should perhaps be emphasised that it will be assumed that carry-over effects depend only on the treatment in the preceding period and not on the treatment in the period being affected by the carry-over. Thus the sequences AA and AB will experience an identical carry-over effect of α_A in the second period.

As is emphasised by Hüsler & Lienert's name for the design, data analysis is facilitated by considering the difference $D_{ij} = y_{ij1} - y_{ij2}$ for subjects in each of the four treatment sequences. By taking this difference, the individual subject effects ω_{ij} are eliminated, as well as the mean μ and sequence effects γ_i. Defining sequence 1 to be AA, sequence 2 BB, sequence 3 AB, and sequence 4 BA, the expressions for the expected value of the differences, are:

$$
\begin{aligned}
E\{D_{1j}\} &= (\pi_1 - \pi_2) - \alpha_A \\
E\{D_{2j}\} &= (\pi_1 - \pi_2) - \alpha_B \\
E\{D_{3j}\} &= (\pi_1 - \pi_2) + (\tau_A - \tau_B) - \alpha_A \\
E\{D_{4j}\} &= (\pi_1 - \pi_2) - (\tau_A - \tau_B) - \alpha_B \qquad .. \ 4.3.1
\end{aligned}
$$

Defining $\bar{D}_{i.}$ to be the average of the n_i D_{ij} values, it is clear that $E\{\bar{D}_{3.} - \bar{D}_{1.}\} = \tau_A - \tau_B$, and $E\{\bar{D}_{4.} - \bar{D}_{2.}\} = -(\tau_A - \tau_B)$, while $E\{\bar{D}_{2.} - \bar{D}_{1.}\} = \alpha_A - \alpha_B$. All of these estimators are of the same form, and hence have similar variances, the variance of $\bar{D}_{i.} - \bar{D}_{j.}$ being

$2\sigma_d^2(1/n_1 + 1/n_3)$.

It would be natural to combine the two estimators of $\tau_A - \tau_B$ in an attempt to obtain an estimator with a smaller variance, by forming a weighted average of the two estimators, with weights inversely proportional to their variances. This yields the estimator

$$C = \{(1/n_2 + 1/n_4)[\bar{D}_{3.} - \bar{D}_{1.}] + (1/n_1 + 1/n_3)[\bar{D}_{2.} - \bar{D}_{4.}]\}/c$$

where $c = 1/n_1 + 1/n_2 + 1/n_3 + 1/n_4$, and the variance is

$$\text{var}(C) = \{2\sigma_d^2(1/n_2 + 1/n_4)(1/n_1 + 1/n_3)\}/c$$

It is interesting to examine how many subjects should be allocated to the four sequences to minimise the variance. It might be expected that equal allocation to each of the four sequences would achieve the minimum variance. If n subjects are allocated to each sequence (i.e. $n_1 = n_2 = n_3 = n_4 = n$) the variance reduces to $2\sigma_d^2(1/n)$. In terms of the total number of subjects, N (=4n), the multiplier of the $2\sigma_d^2$ term is 4/N. If the same N subjects were allocated equally to sequences 1 (AA) and 3 (AB) only, i.e. N/2 subjects to each of these two sequences, and none at all to sequences 2 and 4, the estimator of $\tau_A - \tau_B$ would be $\bar{D}_{3.} - \bar{D}_{1.}$, and the multiplier of the $2\sigma_d^2$ would again be 4/N. As the estimator from sequences 2 (BB) and 4 (BA) has the same variance, this variance could also be achieved by allocating N/2 subjects to sequences 2 and 4, and none to sequences 1 and 3.

It is interesting that the allocation of subjects considered above gives estimators with the same variance as equal allocation to all four sequences even though the number of subjects allocated to complementary sequences (AA and BB, AB and BA), are not the same. In fact, if the N (=4n) subjects are allocated subject only to the restriction that the complementary sequences receive equal numbers of subjects, i.e. m subjects on each of the sequences AA & BB, ($n_1 = n_2 = m$) and 2n-m on each of the sequences AB and BA, ($n_3 = n_4 = 2n-m$) the variance of C is $2\sigma_\varepsilon^2 n/\{m(2n-m)\}$ which is always larger than $2\sigma_\varepsilon^2(4/N)$ unless m=n.

If the number of subjects in complementary sequences is made unequal, with n_1 (AA) = m, n_2 (BB) = 2n-m, n_3 (AB) = m, n_4 (BA) = 2n-m, the variance of C is again $2\sigma_\varepsilon^2(4/N)$. The minimum variance is thus achieved by having equal numbers of subjects in sequences 1&3, and 2&4, but, with this restriction, it does not matter how the subjects are split between these two sets of sequences. It thus appears that a design consisting of the two sequences AA and AB would be a useful design, as it will give a within subject comparison for the treatment difference $\tau_A - \tau_B$. The design will not allow an estimate of the difference in carry-over, but this is only a problem if carry-over is of interest in its own right, as the estimator of the treatment difference is not affected by carry-over. The design may seem rather odd, because of its unbalanced

nature, but if treatment A is a standard treatment and treatment B a new treatment, the design reflects the options open to the physician - continue with the standard, or change to the new. Although this design looks interesting and relatively efficient, it is in fact very similar to a standard parallel design, with a "run-in" period where both sets of subjects are treated with the standard treatment A.

4.3.1 Baseline Observations

It was noted in 4.2 that the addition of baseline observations to the simple two-period cross-over had a beneficial effect on the design, and it is thus of interest to investigate whether baselines have a similar beneficial effect on the complete two-period design, or on the AA, AB design which is derived from it. As before, the linear model will be modified by the introduction of a "period zero" effect, π_0, and expressions for the baseline observations written down. As the combination $E_{ij} = 2y_{ij0} - y_{ij1} - y_{ij2}$ was useful when baselines were present in the simple two-period design, consideration of this combination for each of the four sequences may be beneficial. The expected value of the combination in the four sequences is given below:

$$E(E_{1j}) = (2\pi_0 - \pi_1 - \pi_2) - 2\tau_A - \alpha_A$$

$$E(E_{2j}) = (2\pi_0 - \pi_1 - \pi_2) - 2\tau_B - \alpha_B$$

$$E\{E_{3j}\} = (2\pi_0 - \pi_1 - \pi_2) - \alpha_A$$

$$E\{E_{4j}\} = (2\pi_0 - \pi_1 - \pi_2) - \alpha_B \qquad 4.3.1.1$$

Defining $\bar{E}_{i.}$ as the average of the n_i E_{ij} values, it is clear that $\bar{E}_{3.} - \bar{E}_{1.}$ has expected value $2\tau_A$ and $\bar{E}_{4.} - \bar{E}_{2.}$ has expected value $2\tau_B$. Since τ_A and τ_B are fixed effects with the restriction $\tau_A + \tau_B = 0$, $2\tau_A$ and $-2\tau_B$ are equivalent to the difference $\tau_A - \tau_B$. Thus, two new estimators of the treatment difference can be obtained by using the baseline observations. However, it should be noted that the variance of an E_{ij} combination is $6\sigma_e^2$, as against $2\sigma_e^2$ for a D_{ij} combination, so that an estimate of the treatment difference based on the D's will have a smaller variance than one based on the E's. However, if baseline observations are available, two unbiased estimators of the treatment difference, based on the D's and E's respectively, are possible so that it would make sense to combine the estimates. A consideration of the co-efficients in the D and E combinations shows that they are orthogonal contrasts, since the multipliers of the period 0,1 and 2 observations are 0,1,-1 and 2,-1,-1 respectively. Thus the combination can be expected to produce an estimator with reduced variance, unlike the situation with the simple two-period design. It can be seen from the variances above that the estimator based on the E's has a variance which is three times as large as that based on the D's, so in combining the two estimates

¾(D-estimate) + ¼(E-estimate) will be used. Since both the D's and the E's are linear combinations of the observations, it is not surprising that the combination of the two estimators can be simplified, to a third linear combination $G_{ij} = y_{ij0} + y_{ij1} - 2y_{ij2}$. The expected value of this combination for each of the four sequences is as follows:

$$E\{G_{1j}\} = (\pi_0 + \pi_1 - 2\pi_2) - \tau_A - 2\alpha_A$$
$$E\{G_{2j}\} = (\pi_0 + \pi_1 - 2\pi_2) - \tau_B - 2\alpha_B$$
$$E\{G_{3j}\} = (\pi_0 + \pi_1 - 2\pi_2) + (\tau_A - 2\tau_B) - 2\alpha_A$$
$$E\{G_{4j}\} = (\pi_0 + \pi_1 - 2\pi_2) + (\tau_B - 2\tau_A) - 2\alpha_B \quad .. \ 4.3.1.2$$

G_{ij} is clearly a within-subject comparison and has a variance of $6\sigma_\varepsilon^2$. Defining $\bar{G}_{i.}$ to be the mean of the n_i G_{ij} values, $\frac{1}{2}(\bar{G}_{3.} - \bar{G}_{1.})$ and $\frac{1}{2}(\bar{G}_{2.} - \bar{G}_{4.})$ are unbiased estimators of the treatment difference $\tau_A - \tau_B$, with variances that depend on the number of subjects in each sequence in exactly the same way as the previous estimators based on the D's in 4.2. Hence, as with the combination of the D-estimators, the combination of the two G-estimators will have minimum variance when $n_1 = n_3$, and $n_2 = n_4$. Once again, this optimum is reached if only the sequences AA and AB are used. If there are N/4 subjects in each of the four sequences, or N/2 in sequences 1 and 3 and none in sequences 2 and 4, the variance of the estimator will be $(6/N)\sigma_\varepsilon^2$, compared to

$(8/N)\sigma_\varepsilon^2$ for the estimator based on the D's.

It is of course possible to test the hypothesis that $\tau_A = \tau_B$ by comparing the estimate of the difference with the estimate of the variance of the difference via a t-test.

4.3.2 Washout Observations

As with the simple two-period design, it would be possible to insert a washout period between the first and second treatment periods, and to take a washout observation before the second treatment begins. It is a simple matter to extend the model used for the simple two-period design to the complete two-period design, with the washout period being designated period 1', carry-over effects α_A, α_B affecting the washout observations, and β_A, β_B affecting the observations for the second period. As with the simple two-period design, the washout observations are of little use apart from allowing the estimation of $\alpha_A - \alpha_B$, which is unlikely to be of interest, except as an indication that carry-over is present in the experiment. The estimation of $\tau_A - \tau_B$ is affected by the β's.

Except for ensuring the absence of pharmacological carry-over, and providing a way of checking whether carry-over is present in the experiment by testing for first-order carry-over (α's), there seems little point in taking washout observations.

4.4 Review and Preview

It has been seen that the addition of a single baseline observation, before the first treatment begins, is extremely beneficial in two-period cross-over designs. This fact does not seem to have been widely publicised, and many authors have concentrated instead on the situation where a "baseline" observation is taken before each treatment period (i.e. a combination of "baseline and "washout" observations in the terminology used here). This strategy has been seen to be less helpful, and possibly even detrimental, as it makes the interpretation and modelling of carry-over more awkward. The "complete" two-period cross-over has also received relatively little attention in the literature, even though it is simple, and a marked improvement on the standard two-period cross-over. The author is not aware of any literature on the two sequence AA AB design derived from this "complete" design, even though this appears to be useful. Certainly, if a standard and new treatment are to be compared, this design seems a likely candidate, designating the standard treatment as treatment A. It could be argued that the design is effectively a parallel design with a run-in period, although this has less force if the first period is preceded by a baseline observation. In any case by treating it as a cross-over design, the analysis will almost inevitably be different

from that which would be applied to a parallel design. and arguably, more informative.

In the next chapter the two three-period designs which appear to be useful will be considered in a similar way to the above.

Chapter 5: Classical Analysis for Continuous Response Variables: Three-Period Designs

5.1 Introduction

Analysis of the two two-period designs of interest was considered in chapter 4, including consideration of the effects of baseline and washout observations on the analysis. In this chapter, the analysis of certain three-period designs is considered.

5.2 The Two Sequence Design Without Baselines

In this section the design with two sequences ABB and BAA will be considered when baseline observations are not available. It has already been noted in section 3.3.1 that Kershner & Federer[1981] have shown that this design is efficient in estimating the difference in both the treatment and first-order carry-over effects. The design has also been considered by Ebbutt[1984] and Morrey[1984]. The linear model with parameters as before will be employed, i.e. μ = overall mean, γ_i = sequence effect, ω_{ij} = subject within sequence effect, π_k = period effect, τ_L = treatment effect, α_L = carry-over effect, and ϵ_{ijk} = error term. ω and ϵ are random effects with variances σ_ω^2 and σ_ϵ^2 respectively, while all the others are fixed effects with restrictions as before. Ebbutt[1984] uses essentially the same model, although without specific sequence effects, but gives expressions

for the estimates of treatment and carry-over differences in terms of the quantities T_L, R_L and S_L, which are respectively, the sum of observations on treatment L, the sum of observations in periods preceded by treatment L, and the sum of all observations for those subjects who receive treatment L in the last period. It seems simpler to write down the linear model for each observation with a different combination of sequence and period, and show how linear combinations of these can be combined to obtain the required estimates.

Hence, using the linear model, the following expressions may be obtained:

for subject i in sequence 1;

period 1: $y_{i11} = \mu + \gamma_1 + \omega_{i1} + \pi_1 + \tau_A + \epsilon_{i11}$

period 2: $y_{i12} = \mu + \gamma_1 + \omega_{i1} + \pi_2 + \tau_B + \alpha_A + \epsilon_{i12}$

period 3: $y_{i13} = \mu + \gamma_1 + \omega_{i1} + \pi_3 + \tau_B + \alpha_B + \epsilon_{i13}$

for subject j in sequence 2;

period 1: $y_{2j1} = \mu + \gamma_2 + \omega_{2j} + \pi_1 + \tau_B + \epsilon_{2j1}$

period 2: $y_{2j2} = \mu + \gamma_2 + \omega_{2j} + \pi_2 + \tau_A + \alpha_B + \epsilon_{2j2}$

period 3: $y_{2j3} = \mu + \gamma_2 + \omega_{2j} + \pi_3 + \tau_A + \alpha_A + \epsilon_{2j3}$ ..5.2.1

It should be emphasised that the error terms in the above model are assumed to be independent, so that, given the model parameters, the observations are independent. Because the models for observations on the same subject

in different periods contain the same subject effect, these observations will nevertheless be positively correlated, and the unconditional likelihood will contain covariance terms. If the errors cannot be regarded as independent, a multivariate treatment of the data will be necessary.

The within-subject comparison $H_{ij} = 2y_{ij1} - y_{ij2} - y_{ij3}$ has expected value $(2\pi_1 - \pi_2 - \pi_3) + 2(\tau_A - \tau_B)$ for a subject in sequence 1, and $(2\pi_1 - \pi_2 - \pi_3) - 2(\tau_A - \tau_B)$ for a subject in sequence 2. Thus a two-sample test comparing the H_{1i} and H_{2j} values gives a test of $\tau_A = \tau_B$. An estimate of the treatment difference can be obtained by defining $\bar{H}_{i.}$ as the average of the n_i H_{ij} values and taking $\frac{1}{4}(\bar{H}_{1.} - \bar{H}_{2.})$, giving an unbiased estimate of $\tau_A - \tau_B$ which has variance $\frac{1}{8}\sigma_e^2(1/n_1 + 1/n_2)$.

For estimating or testing the difference in carry-over effects, the within-subject comparison $K_{ij} = y_{ij2} - y_{ij3}$ can be used. For a subject in sequence 1 this has expectation $(\pi_2 - \pi_3) + (\alpha_A - \alpha_B)$, while for a subject in sequence 2 the expectation is $(\pi_2 - \pi_3) - (\alpha_A - \alpha_B)$. Thus, a two-sample test comparing the K_{1i} and the K_{2j} will test for equality of carry-over effects, and, if $\bar{K}_{i.}$ is the average of the n_i K_{ij} values, $\frac{1}{2}(\bar{K}_{1.} - \bar{K}_{2.})$ gives an unbiased estimate of $\alpha_A - \alpha_B$ with variance $\frac{1}{2}\sigma_e^2(1/n_1 + 1/n_2)$.

Ebbutt makes two criticisms of the ABB BAA design. Firstly, he states that the estimators of treatment

difference and first-order carry-over are only orthogonal if there are equal numbers of subjects in the two sequences. This is incorrect, as can be seen from a consideration of the comparisons within a single subject used to obtain the two estimates. The estimator of the treatment difference uses $H_{ij} = 2y_{ij1} - y_{ij2} - y_{ij3}$, while the estimator of carry-over uses $K_{ij} = y_{ij2} - y_{ij3}$. the coefficients of the three observations in these two contrasts are 2, -1, -1 for H and 0, 1, -1 for K, so that the two contrasts are orthogonal. Since the contrasts within each subject are orthogonal, the estimators based on them will be orthogonal, no matter how many subjects are in each sequence.

Ebbutt's second criticism is that blindness may be difficult to maintain if it is known that the treatments in the last two periods are the same. Although this obviously has some force, it would seem possible to maintain blindness, if only by the subterfuge of pretending that there are three treatments involved in the trial. A further problem is the possibility of second-order carry-over effects. If the effect of the treatment administered in the first period carries over to the third period, second-order carry-over effects β_A, β_B, must be introduced into the model. These will appear only in the expressions for the third period observations, and will then be as much of a nuisance in this design as first-order carry-over is in the simple

two-period cross-over. With the addition of these second-order carry-over effects, the expected value of the estimator of the treatment difference $\frac{1}{4}(\bar{H}_{1.} - \bar{H}_{2.})$ is $(\tau_A - \tau_B) - \frac{1}{4}(\beta_A - \beta_B)$; and the expected value of the estimator of the difference in first-order carry-over $\frac{1}{2}(\bar{K}_{1.} - \bar{K}_{2.})$ is $(\alpha_A - \alpha_B) - \frac{1}{2}(\beta_A - \beta_B)$. In addition, it is not possible to test for difference in second-order carry-over effects using a within-subject comparison. The combination $2y_{ij1} + y_{ij2} + y_{ij3}$ must be used, which has expectation $4\mu + 4\gamma_1 + (2\pi_1 + \pi_2 + \pi_3) + \beta_A$ for a subject in sequence 1, and $4\mu + 4\gamma_2 + (2\pi_1 + \pi_2 + \pi_3) + \beta_B$ for a subject in sequence 2. Thus a comparison of the two sets of combinations will only test the hypothesis $\beta_A = \beta_B$ if there is no systematic difference between the subjects in the two sequences (i.e. $\gamma_1 = \gamma_2$); and even if this is the case the test will be inefficient because the subject variation is still present in these combinations. Thus the introduction of second-order carry-over effects into the model has re-introduced all the problems that were present in the simple two-period cross-over.

At this point, it is worth considering whether second-order carry-over effects are really necessary in the model. As mentioned in chapter 2, there are two main ways of viewing carry-over effects. The simplistic view is that carry-over is some pharmacological or psychological effect of the previous treatment that still affects the patient in subsequent periods. Taking this

view, it is reasonable to believe that the carry-over effect will diminish with time, and that second-order carry-over effects will be some fraction of the first-order carry-over effects, so that if there is no difference in first-order carry-over effects, there will not be a difference in second-order carry-over either.

The more complex view of carry-over is that it is some treatment-period interaction. If this view is taken, the "carry-over" can be modelled in the standard way that an interaction is entered into a linear model, rather than introducing carry-over effects. However, in the simple two-period cross-over, first-order carry-over and treatment-period interaction will be aliased and indistinguishable. Similarly, in the three-period design first and second-order carry-over together are equivalent to treatment-period interaction, taking up the same two degrees of freedom in the model. Thus, taking this more complex view of carry-over, first-order carry-over is a component of the total treatment-period interaction. It is difficult to believe that, if some non-specific treatment-period interaction exists its effect would appear solely as second-order carry-over, so that it is reasonable to assume that if there is no evidence of difference in first-order carry-over, there is no treatment-period interaction at all, and hence no difference in second-order carry-over effects. Consequently, whatever view of carry-over is taken, it

seems reasonable to omit second-order carry-over from the model if there is no evidence of first-order carry-over. In addition, the within subject comparison $K_{ij} = y_{ij2} - y_{ij3}$ which was proposed for testing difference in first-order carry-over effects is also affected by any difference in second-order carry-over effects, the expected value of the estimator $\frac{1}{2}(\bar{K}_{1.} - \bar{K}_{2.})$ being $(\alpha_A - \alpha_B) - \frac{1}{2}(\beta_A - \beta_B)$. Thus if the differences in first and second-order carry-over effects have the same sign, which will be the case if second-order carry-over is a diminished form of first-order carry-over, the presence of a non-zero difference in second-order carry-over effects reduces the estimate of the first-order carry-over effect, making it more difficult to detect. Nevertheless, the test of first-order carry-over is quite powerful, and an experimenter may feel that, if this test is non-significant, there will be no need for second-order carry-over effects in the model. However, the corollary to this is that if the test is significant, an experimenter should be reluctant to omit second-order carry-over effects from the model.

Although it has been argued above that significant second-order carry-over is unlikely in the absence of first-order carry-over, the consequences for the analysis of the various combinations of significant and non-significant carry-over effects should be considered. The design is only affected by second order carry-over,

because first-order carry-over is allowed for and does not affect the estimation or testing. Thus, if there is only a difference in the first-order carry-over effects, and not in the second-order carry-over effects, all the data can be retained, and an efficient estimate of treatment difference obtained. If there was a difference in the second-order carry-over effects, but no difference in first-order carry-over effects, the data from the third treatment period would have to be discarded, leaving a simple two-period cross-over design. In this case, significant first-order carry-over would make it necessary to discard the data from the second treatment period also, leaving only a parallel design. Hence, an experimenter who took the conservative approach, and regarded a significant result for the test of first-order carry-over as indicative that there is a carry-over "problem" in the experiment, and that second-order carry-over may well be present also, even if the test for it is non-significant, would have to discard the data from the second and third periods, reducing to a parallel design. Although this might be a sound approach, it does seem rather drastic, since, with carry-over from pharmacological or psychological effects, it is quite possible that the carry-over effects would be strong in the period immediately following a treatment, but would have weakened to become negligible by the second period after the treatment, giving presence of first-order

carry-over differences but absence of difference in second-order carry-over. In these circumstances, the design is very efficient, and indeed has been shown to be optimal for the estimation of treatment and first-order carry-over effects. (Laska et al. [1983])

5.2.1 The Two-sequence Design with Baselines

It was noted in section 4.2 that the addition of baseline observations to the simple two-period cross-over helped to overcome the problems caused by carry-over, and it can be shown that baseline observations will alleviate the similar problems caused by second-order carry-over in the three-period design. Introducing π_0 for the period effect of a baseline observation as before, the model for baseline observations will be exactly as given above for the simple two-period cross-over, i.e.

$$y_{i1o} = \mu + \gamma_1 + \omega_{i1} + \pi_0 + \epsilon_{i1o}$$

$$y_{2jo} = \mu + \gamma_2 + \omega_{2j} + \pi_0 + \epsilon_{2jo}$$

The addition of the baseline observations allows a test of second-order carry-over by a within-subject comparison. The combination $4y_{ijo} - 2y_{ij1} - y_{ij2} - y_{ij3}$ has expected value $(4\pi_0 - 2\pi_1 - \pi_2 - \pi_3) - \beta_A$ for a subject in sequence 1 and $(4\pi_0 - 2\pi_1 - \pi_2 - \pi_3) - \beta_B$ for a subject in sequence 2. Thus a two-sample test comparing the two sets of combinations will test $\beta_A = \beta_B$, and the difference in the means of the two sets of combinations will give an unbiased estimate of $\beta_A - \beta_B$ which has a variance of

$22\sigma^2\{1/n_1 + 1/n_2\}$.

The other possibility presented by baseline observations is for the combination of estimators of treatment difference if carry-over is not present. In these circumstances both the estimator from comparison of the baseline and first period observations, and the cross-over estimator (in this case from the combination $H_{ij} = 2y_{ij1} - y_{ij2} - y_{ij3}$) are valid. It was noted in section 4.2 that combining such estimators for the simple two-period cross-over did not yield a better estimator, because of the positive correlation between the two estimators, and this is again the case for this design.

5.2.2 Baseline and Washout Observations

The analysis of the design will now be considered if there is both a baseline observation before the first treatment period , and washout observations between the first & second, and second & third treatment periods. There are now effectively six periods in the experiment, and as before, the period effect for the baseline observation will be labelled π_0, with the period effects for the washout period between treatment periods 1 & 2, and 2 & 3 being labelled $\pi_{1'}$ and $\pi_{2'}$ respectively. It will be assumed that the six period effects are fixed and sum to zero. First-order carry-over effects α_A, α_B will apply to effects of a treatment carrying over to the following washout period, while second-order carry-over effects β_A,

β_B will apply to treatment effects carrying over from one treatment period to the next, through the intervening washout period. It would be possible to include third and fourth-order carry-over effects, for effects from the first treatment period which affect the second wasout, and third treatment periods respectively, but this seems unnecessarily complicated. With the other model parameters as before, the models for the six observations in each sequence are:

for subject i in sequence 1;

baseline: $y_{1i0} = \mu + \gamma_1 + \omega_{1i} + \pi_0 + \epsilon_{1i0}$

period 1: $y_{1i1} = \mu + \gamma_1 + \omega_{1i} + \pi_1 + \tau_A + \epsilon_{1i1}$

washout : $y_{1i1'} = \mu + \gamma_1 + \omega_{1i} + \pi_{1'} + \alpha_A + \epsilon_{1i1'}$

period 2: $y_{1i2} = \mu + \gamma_1 + \omega_{1i} + \pi_2 + \tau_B + \beta_A + \epsilon_{1i2}$

washout : $y_{1i2'} = \mu + \gamma_1 + \omega_{1i} + \pi_{2'} + \alpha_B + \epsilon_{1i2'}$

period 3: $y_{1i3} = \mu + \gamma_1 + \omega_{1i} + \pi_3 + \tau_B + \beta_B + \epsilon_{1i3}$

for subject j in sequence 2;

baseline: $y_{2j0} = \mu + \gamma_2 + \omega_{2j} + \pi_0 + \epsilon_{2j0}$

period 1: $y_{2j1} = \mu + \gamma_2 + \omega_{2j} + \pi_1 + \tau_B + \epsilon_{2j1}$

washout : $y_{2j1'} = \mu + \gamma_2 + \omega_{2j} + \pi_{1'} + \alpha_B + \epsilon_{2j1'}$

period 2: $y_{2j2} = \mu + \gamma_2 + \omega_{2j} + \pi_2 + \tau_A + \beta_B + \epsilon_{2j2}$

washout : $y_{2j2'} = \mu + \gamma_2 + \omega_{2j} + \pi_{2'} + \alpha_A + \epsilon_{2j2'}$

period 3: $y_{2j3} = \mu + \gamma_2 + \omega_{2j} + \pi_3 + \tau_A + \beta_A + \epsilon_{2j3}$

..5.2.2.1

The advantage of inserting washout periods can be seen from the fact that first and second-order carry-over effects have been separated out, and do not occur together in any of the above expressions. This enables within-subject comparisons which will test for differences in treatment, first and second-order carry-over effects. The familiar combination of the three treatment periods $H_{ij} = 2y_{ij1} - y_{ij2} - y_{ij3}$, will provide a test of differences between the treatments. First-order carry-over effects are not involved in this combination, and second-order effects are eliminated from its expected value by the restriction $\beta_A + \beta_B = 0$. Hence, in the absence of higher-order carry-over effects, carry-over does not affect the test or estimation. Similarly, the comparison of the two washout observations $D'_{ij} = y_{ij1'} - y_{ij2'}$, gives a test of equality of first-order carry-over effects without involving second-order effects; and the difference between the observations for the second and third treatment periods $K_{ij} = y_{ij2} - y_{ij3}$, gives a test of the equality of the second-order carry-over effects. It is interesting to note that the baseline observations are not invoved in any of these comparisons, suggesting that the baseline observations are redundant and need not be taken. Comparison of the baseline observations would be used to test for a systematic difference between the two groups of patients (i.e. $\gamma_1 = \gamma_2$), which might still be regarded as important, but as

the γ's are eliminated from within-subject comparisons such as those above for testing for treatment, first-order, and second-order carry-over effects, any difference between the groups would not affect the tests for these effects.

The introduction of washout periods seems from the above analysis to have solved all the problems of this cross-over design. However, the problems have only disappeared because of the parameterisation that has been used, and in particular the simplistic view that has been taken of carry-over. By not including higher than second-order carry-over effects we are tending to the pharmacological/psychological view of carry-over, and assuming that such effects will diminish with time, becoming negligible after two periods. This may be reasonable if pharmacological/psychological effects really are involved, but it should be remembered that carry-over is intended to cover any effect that makes a treatment be perceived differently in different treatment periods. Although pharmacological/psychological carry-over is the most obvious mechanism for such a treatment-period interaction, we should not be seduced into thinking that carry-over must arise in this way. On the other hand, if we insist on introducing parameters into the model which allow the treatments to be different in the different periods, we inevitably re-introduce the familiar problems associated with the two-period cross-

over. If it is believed that at least some of any treatment-period interaction will be due to pharmacological/psychological effects, and that these can be reduced by extending the design by introducing washout periods, it will be worth taking these steps,even though other, less-obvious causes of treatment-period interaction may remain.

5.3 Four Sequence Designs

As an alternative to the three-period design with sequences ABB & BAA, Ebbutt[1984] considers the design with the four sequences ABB, BAA, ABA & BAB. Kershner & Federer[1981] also consider this design and two other four sequence three-period designs : AAB, BBA, ABB & BAA; and AAB, BBA, ABA & BAB. Kershner & Federer show that the last of these designs gives estimates of treatment and first-order carry-over effects which have much larger variances than those from the other two designs, and this design will be ignored. In chapter 3, it was suggested that,by analogy with the complete two period design, it might be worth considering the design with sequences AAA, BBB, ABB & BAA. Hence the three four-sequence designs shown in table 3 will be considered in this section.

Table 3: Sequences in the three designs.

Design No.	Sequences			
1	ABB	BAA	ABA	BAB
2	ABB	BAA	AAB	BBA
3	AAA	BBB	ABB	BAA

Clearly, apart for the fact that four sequence effects $\gamma_1 \ldots \gamma_4$ will be required, the parameterisation used for the design ABB, BAA in section 5.2 above can be used to define expressions for each observation in each of the three designs. First and second-order carry-over effects will be included from the outset; Kershner & Federer have identified the best linear unbiased estimators for treatment and first-order carry-over differences if second-order carry-over is omitted from the model. Initially it will be assumed that an equal number of subjects is allocated to each of the four sequences in the designs, and the within-subject comparisons that give the best unbiased estimators of the difference in treatment effects will be identified.

For design 1, the comparison required for each sequence is $L_{ij} = y_{ij1} - y_{ij3}$. This comparison yields the following expected values for each of the four sequences;

$$E\{L_{1j}\} = (\pi_1 - \pi_3) + (\tau_A - \tau_B) - \alpha_B - \beta_A$$

$$E\{L_{2j}\} = (\pi_1 - \pi_3) - (\tau_A - \tau_B) - \alpha_A - \beta_B$$

$$E\{L_{3j}\} = (\pi_1 - \pi_3) - \alpha_B - \beta_A$$

$$E\{L_{4j}\} = (\pi_1 - \pi_3) - \alpha_A - \beta_B \qquad ..5.3.1$$

Thus, writing $\bar{L}_{i.}$ for the average of the L values in sequence i, the combination $\frac{1}{8}(\bar{L}_{1.} - \bar{L}_{2.} - \bar{L}_{3.} + \bar{L}_{4.})$ gives an unbiased estimate of $\tau_A - \tau_B$, with variance $2\sigma_d^2/n$ or $8\sigma_d^2/N$, if there are n subjects in each of the four sequences, and N = 4n subjects altogether.

For design two the comparison $M_{ij} = 5y_{ij1} - 4y_{ij2} - y_{ij3}$ is used for sequences 1 & 2 (ABB & BAA), and $M'_{ij} = 3y_{ij1} - 2y_{ij2} - y_{ij3}$ for sequences 3 & 4 (AAB & BBA). These give the following expected values;

$$\begin{aligned}
E\{M_{1j}\} &= (5\pi_1 - 4\pi_2 - \pi_3) + 5(\tau_A - \tau_B) - 4\alpha_A - \alpha_B - \beta_A \\
E\{M_{2j}\} &= (5\pi_1 - 4\pi_2 - \pi_3) - 5(\tau_A - \tau_B) - \alpha_A - 4\alpha_B - \beta_B \\
E\{M'_{3j}\} &= (3\pi_1 - 2\pi_2 - \pi_3) + (\tau_A - \tau_B) - 3\alpha_A - \beta_A \\
E\{M'_{4j}\} &= (3\pi_1 - 2\pi_2 - \pi_3) - (\tau_A - \tau_B) - 3\alpha_B - \beta_B \qquad .5.3.2
\end{aligned}$$

Assuming there are n = N/4 subjects in each of the four sequences, and writing $\bar{M}_{i.}$ and $\bar{M}'_{i.}$ for the averages of the M and M' combinations, the estimator $\frac{1}{8}(\bar{M}_{1.} - \bar{M}_{2.} - \bar{M}'_{3.} + \bar{M}'_{4.})$ is unbiased for $\tau_A - \tau_B$, and has variance $1.75\sigma_d^2/n$ or $7\sigma_d^2/N$.

For design 3, two different combinations are again used, $P_{ij} = 2y_{ij1} + y_{ij2} - 3y_{ij3}$ for sequences 1 & 2 (AAA & BBB) and $P'_{ij} = 8y_{ij1} - 5y_{ij2} - 3y_{ij3}$ for sequences 3 & 4 (ABB & BAA). These result in the expected values;

$$E\{P_{1j}\} = (2\pi_1 + \pi_2 - 3\pi_3) - 2\alpha_A - 3\beta_A$$

$$E\{P_{2j}\} = (2\pi_1 + \pi_2 - 3\pi_3) - 2\alpha_B - 3\beta_B$$

$$E\{P'_{3j}\} = (8\pi_1 - 5\pi_2 - 3\pi_3) + 8(\tau_A - \tau_B) - 5\alpha_A - 3\alpha_B - 3\beta_A$$

$$E\{P'_{4j}\} = (8\pi_1 - 5\pi_2 - 3\pi_3) - 8(\tau_A - \tau_B) - 3\alpha_A - 5\alpha_B - 3\beta_B$$

..5.3.3

Writing $\bar{P}_{i.}$ and $\bar{P}'_{i.}$ for the averages of the combinations as appropriate, and assuming n = N/4 subjects in each sequence, the combination $(\bar{P}_{1.} - \bar{P}_{2.} - \bar{P}'_{3.} + \bar{P}'_{4.})/16$ is an unbiased estimator of $-(\tau_A - \tau_B)$ with variance $0.875\sigma^2/n$ or $7\sigma^2/2N$.

It is interesting that the estimator with the smallest variance arises from the design which is a combination of the cross-over and the parallel designs, confirming the apparent usefulness of this strategy, which was noted with the two-period designs. Since unbiased estimators of the treatment difference have been obtained from these designs, no attempt will be made to consider the addition of baseline and washout observations. The estimators obtained are efficient, in the sense that the model used in obtaining them contains both first and second-order carry-over effects, these being equivalent to treatment-period interaction for a three period design.

The variances given for the estimates have assumed that n = N/4 subjects were allocated to each of the four sequences, but it may be that the variances can be

reduced by unequal allocation. It should be noted at the outset that observations on all four sequences are required for each design, unlike the two-period case where the two sequences AA & AB were found to be as efficient as the four sequences AA, BB, AB & BA. Consideration of design 3 above shows that the variance of the estimate of treatment difference can be reduced by keeping the same number of subjects in the complementary sequences (AAA & BBB, and ABB & BAA) but by having unequal numbers in these two sequence types. If m subjects are allocated to the ABB & BAA sequences, and 2n-m subjects to the AAA & BBB sequences, so that a total of 4n = N subjects take part in the experiment, the variance of the estimator of $\tau_A - \tau_B$ is $\sigma_i^2\{98(2/m) + 14(2/[2n-m])\}/256$, which reduces to $7\sigma_i^2\{7/m + 1/(2n-m)\}/64$. It can be shown that the value of m which minimises this variance is $(7 - 7^*)n/3$, making the ratio of the number of subjects in sequences 3 (ABB) and 4 (BAA) to the number in sequences 1 (AAA) and 2 (BBB) 7^* : 1. With a total of 4n subjects in the four sequences, this then makes the variance of the estimator equal to $\{(7^* + 4)7\sigma_i^2\}/64n = 0.7269\sigma_i^2/n = 2.907\sigma_i^2/N$. Of course, it is not possible to make the optimum allocation of subjects with a ratio of 7^* : 1, but this can be approximated to. If the ratio of the number of subjects in sequences 3 & 4 to the number in sequences 1 & 2 is 5 : 2 or 8 : 3, the variance of the estimate of treatment

difference is approximately $2.91\sigma_\varepsilon^2/N$. Similar improvements could be made for the other two designs, but these would not yield an estimator with a lower variance.

5.4 Review

In chapter 4 it was found that the addition of baseline observations to a two-period cross-over design lead to considerable improvement in the estimates that are possible, but that washout observations between the two treatment periods were comparatively unhelpful. In the present chapter, the usefulness of baseline observations with three-period designs has again been apparent, but washout observations have also been found to be useful. In a two-period design without washout observations, there is no place for second-order carry-over in the model, so the introduction of these effects with wash-out observations creates a further complication. On the other hand, second-order carry-over will affect the third treatment period in a three-period design, in addition to first-order carry-over. The addition of washout observations in this case separates first and second-order carry-over leading to a simplification of the analysis.

Chapter 6: The two-period Cross-over with Binary Data

6.1 Introduction

In clinical trials, it is not uncommon for the outcome to be a dichotomy, for example improved/not improved or relief/no relief. In general, the two mutually exclusive and exhaustive outcome classifications can be thought of as success and failure for the treatment. Such results are usually recorded as 0 for failure and 1 for success, giving rise to what is termed binary data. Many authors have considered the analysis of such binary data for the simple two-period cross-over, although many of the tests given have been based on intuition and common sense, rather than any formal mathematical modelling of the data. Kenward & Jones[1987a] have attempted to put these tests on a formal basis by relating them to a general model for binary data, but have also generated some controversy as to the most appropriate way of modelling such data. This chapter reviews the various tests proposed and discusses the modelling strategies, in preparation for consideration of the analysis of more complex cross-over designs in chapter 7.

6.2 Models for Binary Data

In considering models for binary data, the notation y_{ijk} = response for subject j in sequence i for period k

will be retained. However, this response is now binary, typically 0 (failure) or 1 (success). Interest will now be in the probability of a particular treatment giving a successful outcome, but this is, of course, not directly observable. The probability that a particular observation is "success" i.e. $p(y_{ijk} = 1)$, will depend not only on the treatment that was applied, but also on the subject and other circumstances such as the period. Following the usual practice with continuous observations, it would be natural to use a linear model containing parameters for these various effects to model $p(y_{ijk} = 1)$. Such an approach has been used for the two-period cross-over by Zimmerman & Rahlfs[1978], and will be considered in more detail later.

The problem with modelling the probability directly is that the linear model must then yield a value between 0 and 1, which is difficult to ensure. To avoid this difficulty, it is common practice to transform the probability into a variable which can take any value between $-\infty$ and ∞, before employing a linear model. The functions most commonly used for this purpose are the probit and logit transformations, both of which are discussed in Cox[1970]. Most of the tests proposed for the simple two-period cross-over relate to the logit transformation, while Cox & Plackett[1980] have proposed a test based on the probit transformation. Since the two transformations have similar characteristics, there seems

to be little to choose between them, although logit models may be slightly more convenient to use.

If a linear model for the logit, or probit, of the probability of success is to be used, it would seem reasonable to make it identical in form with that which was used for continuous responses. Thus, writing $\text{logit}\{p(y_{ijk} = 1)\} = \lambda_{ijk}$, we have, for instance:

$$\lambda_{ij2} = \mu + \gamma_i + \omega_{ij} + \pi_2 + \tau_B + \alpha_A \quad ..6.2.1$$

where γ is the group or sequence effect, ω is the subject effect, π is the period effect, τ is the treatment effect, and α is the carry-over effect from the previous treatment.

There are several problems with the use of such a model for the binary data from a cross-over trial. For the simple two-period cross-over it will be found that the sequence and carry-over effects are aliased, as in the case of continuous responses. In fact the estimation and testing problems that have been discussed for the case of continuous responses will continue to complicate the analysis for binary data. With binary observations, there is perhaps more justification for simplifying the model than with continuous observations. The multiplicity of parameters can help to explain slight differences in observations with continuous responses, but slight differences in the logit or probit, and hence probability, will be hidden by the binary nature of the observed response. The subject parameters are a

particular problem, as very little information can be gained about them. However, it is intuitively reasonable to believe that subjects will react differently in similar circumstances. In terms of the observed binary response, this will mean that some subjects are more likely than others to give a "success ", which will in turn imply a correlation between the observations on the same subject at different times.

Because of the correlation between the two responses on each subject in the simple two-period cross-over, some authors have considered the two responses for a subject as the basic observation. There are then four different possible patterns of response that can be observed: 00, 01, 10 and 11,and the number of times each pattern is observed for subjects receiving a particular treatment sequence can be thought of as having a multinomial distribution. Taking this view, it is then possible to construct a linear model for the four probabilities associated with the four different patterns, rather than for the probability of success for an individual response. If the four possible responses for a single subject are considered as forming a 2x2 table, the probabilities associated with the marginal events are those considered originally, i.e. $p(y_{ijk} = 0)$ or $p(y_{ijk} = 1)$.

		period 2 outcome		
period		failure (0)	success (1)	
1	failure (0)	00	01	$p(y_{ij1} = 0)$
outcome	success (1)	10	11	$p(y_{ij1} = 1)$
		$p(y_{ij2} = 0)$	$p(y_{ij2} = 1)$	

The advantage of modelling these marginal probabilities, or a function of them, is the ease of interpretation of the period, treatment, and other parameters, but the disadvantage is that, if subject effects are used these are nuisance parameters. The advantage of modelling the joint probabilities is that, by considering them as the multinomial probabilities, they could be assumed to be averaged over subjects, making individual subject effects irrelevant. The disadvantage is that parameters in a linear model do not have a simple interpretation, or relate in a simple way to the period, treatment, and other effects associated with the marginal or individual probabilities.

The problem of modelling binary data for cross-overs thus revolves around the choice of whether to model the joint or marginal probabilities, and how to model the dependence structure between the observations on the same subject. The most natural choice seems to be to model the marginal probabilities and include subject parameters, as for the case of continuous responses, but this proves to be rather intractable. Kenward & Jones[1987a] have also

modelled the marginal probabilities, but have suggested that the dependence structure should be treated separately, and modelled by adding appropriate parameters to a linear model from which subject effects have been omitted. Although this much simplifies the modelling, the resulting dependence parameters are not related in any obvious way to the subject parameters they replace, and seem rather arbitrary. Fidler[1984,1986] favours modelling the joint probabilities, but this leads to a more complex interpretation of the parameters involved. The various tests for the two-period cross-over, and the models on which they are based will now be briefly considered.

6.3 The Mainland-Gart Test

Historically, the first test to be proposed was the Mainland-Gart test, suggested by Mainland[1963] on common-sense grounds, and given a rigorous foundation by Gart[1969]. Gart used a logistic or log-linear model, in which the log of the odds ratio, or logit, of the probability of success for a particular observation is expressed as a linear model. i.e. for an observation y, $\text{logit}\{p(y = 1)\} = \ln\{p(y = 1)/p(y = 0)\} = L$, where L is the linear model. Thus

$$p(y = 1) = e^L/(1 + e^L)$$

$$p(y = 0) = 1/(1 + e^L)$$

Gart's linear model contained terms for subject, period and treatment effects, but contained no term for carry-over. An overall mean term was also omitted from the model, being effectively included in the subject effects. Using the notation of the previous chapters, the model is;

for subject i in sequence 1:

$$\text{logit}\{p(y_{i1} = 1)\} = \exp(\omega_{i1} + \pi_1 + \tau_A)$$

$$\text{logit}\{p(y_{i2} = 1)\} = \exp(\omega_{i1} + \pi_2 + \tau_B)$$

for subject j in sequence 2:

$$\text{logit}\{p(y_{2j1} = 1)\} = \exp(\omega_{2j} + \pi_1 + \tau_B)$$

$$\text{logit}\{p(y_{2j2} = 1)\} = \exp(\omega_{2j} + \pi_2 + \tau_A) \quad ..6.3.1$$

Gart is thus adapting the usual method for continuous responses to binary data, using subject effects to account for the dependence between the two observations on the same subject. Given the subject effect, the joint probability of a particular pattern of two responses can be obtained from the product of the two individual probabilities. With the usual restrictions $\pi_1 + \pi_2 = \tau_A + \tau_B = 0$, this gives the following probabilities of the four different patterns of responses, (00, 01, 10 & 11).

for subject i in sequence 1:

$p(00) = 1/T_i$

$p(01) = \exp(\omega_{1i} + \pi_2 + \tau_B)/T_i$

$p(10) = \exp(\omega_{1i} + \pi_1 + \tau_A)/T_i$

$p(11) = \exp(2\omega_{1i})/T_i$

for subject j in sequence 2:

$p(00) = 1/T'_j$

$p(01) = \exp(\omega_{2j} + \pi_2 + \tau_A)/T'_j$

$p(10) = \exp(\omega_{2j} + \pi_1 + \tau_B)/T'_j$

$p(11) = \exp(2\omega_{2j})/T'_j$ ..6.3.2

where $T_i = \{1 + \exp(\omega_{1i} + \pi_1 + \tau_A)\}\{1 + \exp(\omega_{1i} + \pi_2 + \tau_B)\}$

and $T'_j = \{1 + \exp(\omega_{2j} + \pi_1 + \tau_B)\}\{1 + \exp(\omega_{2j} + \pi_2 + \tau_A)\}$.

It can be seen from these expressions that only the probabilities for the patterns 01 and 10 contain the parameters of interest (π & τ) in the numerator; that is, apart from the normalising constants T_i and T'_j. This reinforces the common-sense idea that information about the relative merits of the two treatments can only be obtained from subjects who react differently to them. Hence it seems sensible to find the conditional probabilities of success for treatment A, or success in period 1 given that the two responses are different. Note that this is conditioning on the sufficient statistic for the subject effect, which is the number of successes; rather than the more usual practice of conditioning on an

ancillary statistic. It is only when there is one success that there is any choice in the pattern of response, and thus any information about parameters other than the subject effects we are conditioning on. The expressions for the conditional probabilities are:

for subject i in sequence 1:

$$p(\text{success for treatment A} \mid \text{only 1 success}) =$$

$$p(\text{success in period 1} \mid \text{only 1 success}) =$$

$$\{\exp(\pi_1 + \tau_A)\}/\{\exp(\pi_1 + \tau_A) + \exp(\pi_2 + \tau_B)\}$$

for subject j in sequence 2:

$$p(\text{success for treatment A} \mid \text{only 1 success}) =$$

$$\{\exp(\pi_2 + \tau_A)\}/\{\exp(\pi_1 + \tau_B) + \exp(\pi_2 + \tau_A)\}$$

$$p(\text{success in period 1} \mid \text{only 1 success}) =$$

$$\{\exp(\pi_1 + \tau_B)\}/\{\exp(\pi_1 + \tau_B) + \exp(\pi_2 + \tau_A)\} \quad ..6.3.3$$

These expressions no longer contain the subject effects, and it can be seen that p(success in period 1 | only 1 success) is the same for both sequences if $\tau_A = \tau_B$, giving a test of equality of treatment effects; while p(success for treatment A | only 1 success) is the same for the two sequences if $\pi_1 = \pi_2$, giving a test of the equality of the period effects. These tests are made by testing for the equality of these conditional probabilities using the estimates given by the observations. If N subjects receive treatment sequence AB

of which n_{00}, n_{01}, n_{10} and n_{11} give responses 00, 01, 10 & 11 respectively, and N' subjects receive treatment sequence BA with n'_{00}, n'_{01}, n'_{10} and n'_{11} giving responses 00,01,10 & 11, the estimates of p(success in period 1 | only 1 success) are $n_{10}/(n_{01} + n_{10})$ for sequence 1, and $n'_{10}/(n'_{01} + n'_{10})$ for sequence 2. If the treatment effects are equal, the expected values of these estimates are equal, so that the hypothesis of equal treatment effects may be tested by the usual χ^2 test for the contingency table

		seq 1	seq 2
success	1	n_{10}	n'_{10}
in period	2	n_{01}	n'_{01}

Similarly, the estimates of the conditional probabilities p(success for treatment A | only 1 success) are $n_{10}/(n_{01} + n_{10})$ for sequence 1 and $n'_{01}/(n'_{01} + n'_{10})$ for sequence 2. The hypothesis of equal period effects implies that the expected values of these estimates are equal, so that the hypothesis can be tested by the usual χ^2 test for the contingency table

		seq 1	seq 2
success for	A	n_{10}	n'_{01}
treatment	B	n_{01}	n'_{10}

It should be noted that in testing for a difference between treatments, the entries in each row refer to the same period, while in testing for a difference between periods each row refers to a treatment, so that the periods are reversed for sequence 2. It will be found that it is generally true that a test for periods will result when the order of the observations relating to sequence 2 are reversed in a test for treatments.

6.4 Zimmerman & Rahlfs Model

Zimmerman & Rahlfs[1978] model the marginal probabilities directly, without making a logit or probit transformation. Using the usual notation y_{ijk} = response for subject j in sequence i for period k, the model is:

$$
\begin{aligned}
p(y_{1j1} = 1) &= p_{1.} = \mu + \pi_1 + \tau_A \\
p(y_{1j2} = 1) &= p_{.1} = \mu + \pi_2 + \tau_B + \alpha_A \\
p(y_{2j1} = 1) &= q_{1.} = \mu + \pi_1 + \tau_B \\
p(y_{2j2} = 1) &= q_{.1} = \mu + \pi_2 + \tau_A + \alpha_B \qquad ..6.4.1
\end{aligned}
$$

where, as usual, μ is an overall mean, π_1, π_2 are period effects, τ_A, τ_B are treatment effects, and α_A, α_B carry-over effects.

It will be noticed that no subject effects are included in the model, so that the probabilities must be regarded as the average over a large number of subjects, and the correlation between observations on the same

subject is not modelled. Since the model is identical in form to the model for continuous responses, the combinations used for estimation in that case can be applied here. Hence, if carry-over effects are not present, i.e. $\alpha_A = \alpha_B = 0$, we expect $p_{1.} + p_{.1}$ to be equal to $q_{1.} + q_{.1}$, and $p_{1.} + p_{.1} - q_{1.} - q_{.1}$ is an estimator of $\alpha_A - \alpha_B$. Tests can be performed using the methodology proposed by Grizzle, Starmer & Koch[1969], based on the minimum logit χ^2, and Zimmermann & Rahlfs give explicit formulae leading to equivalent χ^2 tests. As with the continuous case, if the test indicates that there is no carry-over, the hypothesis of equal treatment effects leads us to expect that $p_{1.} - p_{.1} = q_{1.} - q_{.1}$, and the hypothesis of equal period effects implies that $p_{1.} - p_{.1} = q_{.1} - q_{1.}$. If there is carry-over present, the test for treatment effects will be based on the responses for the first period in the two different sequences.

6.5 Cox & Plackett's Models

Cox & Plackett[1980] propose two models for the transformed marginal probabilities, one of which uses the logit transformation, and the other the probit transformation. Apart from using different transformations, the two models also differ in form, subject effects being present in the probit model, but not in the logit model. Writing $\lambda_{ijk} = \text{logit}(y_{ijk})$, the logit model is:

$$\lambda_{1j1} = \mu$$

$$\lambda_{1j2} = \mu + \pi + \tau + \alpha$$

$$\lambda_{2j1} = \mu + \tau$$

$$\lambda_{2j2} = \mu + \pi \qquad ..6.5.1$$

where μ is a location constant, π relates to period differences, τ relates to treatment differences, and α relates to treatment x period interaction. It should be noted that this is effectively the same model as used before, but with different constraints on the parameters, i.e. $\pi_1 = \tau_A = 0$, instead of $\pi_1 + \pi_2 = \tau_A + \tau_B = 0$.

This model assumes that the equality of the cross-product ratio $p_{11}p_{00}/p_{10}p_{01} = q_{11}q_{00}/q_{10}q_{01}$ applies, and this is tested using the test of zero interactions for a 2x2x2 table given by Plackett[1974]. If this test is non-significant, the logits and their variances and covariances are estimated from the observed frequencies, to give tests of the individual parameters.

The probit model is:

$$p(y_{1j1} = 1) = \phi(\mu + \omega_{1j})$$

$$p(y_{1j2} = 1) = \phi(\mu + \pi + \tau + \alpha + \omega_{1j})$$

$$p(y_{2j1} = 1) = \phi(\mu + \tau + \omega_{2j})$$

$$p(y_{2j2} = 1) = \phi(\mu + \pi + \omega_{2j}) \qquad ..6.5.2$$

where $\phi(x)$ is the cumulative normal probability, and μ, π, τ and α are as for the previous model.

Testing for this model invoves using numerical methods to maximise the log likelihood, in order to obtain a deviance.

6.6 Fidler's Model

Fidler[1984] models the joint probabilities rather than the marginal probabilities. Thus, in his notation the probability models are:

Outcome	Seq 1 (AB)	Seq 2 (BA)
00	C_1	C_2
01	$C_1\exp\{a_{11} - \tau - \pi\}$	$C_2\exp\{a_{12} + \tau - \pi\}$
10	$C_1\exp\{a_{11} + \tau + \pi\}$	$C_2\exp\{a_{12} - \tau + \pi\}$
11	$C_1\exp\{a_{21}\}$	$C_2\exp\{a_{22}\}$ ..6.6.1

C_1 and C_2 are normalising constants, and τ, π, a_{11}, a_{12}, a_{21} and a_{22} are the model parameters. For each sequence, the probability of the four possible outcomes forms a quadrinomial distribution, so these six model parameters are equivalent to the two sets of three unknown probabilities, and constitute a fully saturated model. τ and π correspond to the main effects of treatments and periods, with τ having a positive coefficient when the success is for treatment A, and π having a positive coefficient when the success is in period one. When there is a success on both treatments, and in both periods, the restriction that the sums of the

treatment effects and of the period effects are zero means that neither τ nor π are required in the model. The parameters a_{11}, a_{12}, a_{21}, a_{22} must control the group and carry-over effects. If there is neither a group nor a treatment effect, the probability of success in period one would be the same for each sequence, however, this marginal probability contains both a_{11} and a_{21} in sequence 1 and a_{12} and a_{22} in sequence 2. Similarly, if there is neither a treatment nor a group effect, the absence of a carry-over effect would imply that the probability of success in the second period was the same for each sequence. Again the marginal probability of success in period 2 will contain a_{11} and a_{21} in sequence 1 and a_{21} and a_{22} in sequence 2. The parameters can thus not be separated out in any simple way to give each a simple interpretation, but if $a_{11} = a_{12} = a_1$, and $a_{21} = a_{22} = a_2$, both group and carry-over effects are absent. By considering the fully saturated model, and models omitting group, group and carry-over, and period effects, Fidler obtains tests for treatments. Using the fully saturated model, the test for treatments obtained is just the Mainland-Gart test, while omitting group and carry-over effects leads to the test first proposed by Prescott[1981], which is discussed in 6.8 below. Tests for carry-over are also obtained by considering departures from the model with no group or carry-over effects of the form $a_{21} - a_{22} \neq 0$. Two tests are possible,

depending on whether $a_{11} - a_{12}$ is considered to be zero. The simpler of these was proposed by Armitage & Hills[1982] and is also discussed in 6.8 below.

Although Fidler's model seems more complicated to interpret than models for the marginal probabilities, he has pointed out in a letter (Fidler[1988]) that parameters in models for marginal probabilities often need to be interpret ed with more care than might at first seem necessary, especially if there are no subject parameters to take care of the dependence structure in the data.

6.7 Kenward and Jones' Model

Kenward and Jones[1987a] model the marginal probabilities with a logit model with treatment, period and carry-over effects, a parameter representing the overall tendency to give a positive response and normalising constants for the two sequences, which are equivalent to the overall mean and group effects. Ignoring the normalising constants, four parameters are therefore included to take care of the effects of interest, which, by analogy with the six parameters available for two quadrinomial distributions, implies that two more parameters are available for a saturated model. These two parameters are used to take care of the dependence structure in the data, one, σ, being described as the "average" within-subject dependence, and the

second, ϕ, as the difference in dependence between the two groups, or the group-by-dependence interaction. With C_1 and C_2 as normalising constants, μ as the overall tendency of subjects to give a success, and π, τ, and α as the period, treatment and carry-over effects respectively, the model for the possible outcomes in sequence 1 are:

$$
\begin{aligned}
&00 \qquad C_1\exp\{\sigma + \phi\} \\
&01 \qquad C_1\exp\{\mu + \pi + \tau - \alpha - \sigma - \phi\} \\
&10 \qquad C_1\exp\{\mu - \pi - \tau - \sigma - \phi\} \\
&11 \qquad C_1\exp\{2\mu - \alpha + \sigma + \phi\}
\end{aligned}
$$

while for sequence 2 they are:

$$
\begin{aligned}
&00 \qquad C_2\exp\{\sigma - \phi\} \\
&01 \qquad C_2\exp\{\mu + \pi - \tau + \alpha - \sigma + \phi\} \\
&10 \qquad C_2\exp\{\mu - \pi + \tau - \sigma + \phi\} \\
&11 \qquad C_2\exp\{2\mu + \alpha + \sigma - \phi\} \qquad\qquad ..6.7.1
\end{aligned}
$$

The inclusion of the parameters σ and ϕ is meant to mimic the situation that would result if the subject parameters were integrated out. However, the authors themselves admit that it is difficult to envisage how such a process could result in such a convenient parameterisation. If subject parameters were integrated out, the marginal probabilities would be independent, so the model is an attempt to get the benefit of having

independent marginal probabilities, while avoiding the nuisance of the subject parameters themselves. If the subject parameters are considered to be random effects with normal distributions, as in the continuous case, integrating them out proves very difficult, and requires numerical integration methods.

Kenward & Jones use their model to evaluate the various tests that have been proposed by identifying the parameters in their model which are involved in each test. In doing so, they are concerned if a test apparently breaches the marginality rules by fitting carry-over (or treatment-period interaction) when testing for treatments or periods (i.e. with one of the main effects for the interaction potentially omitted). The introduction of the idea of treatment-period interaction in place of carry-over was noted earlier as a way of avoiding the simplistic view of carry-over as a mere pharmacological effect, but this concern over marginality seems to be an almost equally mis-placed view of carry-over, as it can reasonably be regarded as a genuine effect of a treatment. In any case, the marginality rules do not seem to be universally agreed upon, as can be seen from the literature associated with the fitting of linear models with the SAS package (see e.g. Pendleton et al[1986])

6.8 Other Tests

6.8.1 Prescott's Test

Although it seems intuitively reasonable that 00 and 11 responses give no information about the relative merits of the two treatments, it is nevertheless uncomfortable to discard this data, and it may sometimes mean that the remaining 2x2 table for Gart's test is sparse. Prescott[1981] reasoned that there are three types of responses: 10, 00 & 11, and 01, that are progressively less favourable to the first treatment, and more favourable to the second. Since the first treatment is different in the two sequences, there will be no difference in the relative frequencies of these responses in the two sequences, provided there is no difference in the treatments, and no carry-over. Alternatively, if there is a difference in the treatments, a trend in the three types of response can be expected, being in opposite directions for the two sequences. Thus a test based on the 2x3 table given below will test for treatment difference.

response:	10	00 or 11	01
seq 1	n_{10}	$n_{00} + n_{11}$	n_{01}
seq 2	n'_{10}	$n'_{00} + n'_{11}$	n'_{01}

The proposed test can be thought of as arising from the regression of dummy variable X, characterising the response type, and dummy variable Y, characterising the sequence (Cox[1970], p61-64). Variable X may conveniently be defined as taking the value 1 if the response is 10, zero if the response is 00 or 11, and -1 if the response is 01; while variable Y takes the value 1 for sequence 1 and 0 for sequence 2. This means that $\Sigma xy = n_{10} - n_{01}$. Given fixed margins in the 2x3 table, the evidence of a relationship between X and Y increases with the magnitude of Σxy. The probability of obtaining a value of Σxy that is as extreme or more extreme than the observed value is calculated using a randomisation test which assumes that the subjects in each of the three response catagories are allocated to the two sequences at random.

6.8.2 Tests of Carry-over

Clearly, as Gart's logistic model contains no term for carry-over, it is not possible to perform a test for carry-over effects under his scheme. However Hills & Armitage[1979, 1982] have proposed two tests of carry-over, both on common-sense grounds. The first of these (Hills & Armitage[1979]) arises by analogy with the situation with continuous response variables.

It will be recalled that testing for carry-over with continuous responses involves comparing the sum of the two observations for subjects in the two sequences. If

there is no carry-over the sums are expected to have the same mean in the two sequences, i.e. the average level of response will be the same. For binary responses, the equivalent consequence is that approximately equal proportions of 0 and 1 responses will be expected in each sequence. Since the number of 0 & 1 responses are most affected by variation in the number of 00 and 11 patterns, this can be tested by comparing the relative frequencies of these in the two sequences; i.e. the association in the 2x2 contingency table:

	seq.1	seq.2
00	n_{00}	n'_{00}
11	n_{11}	n'_{11}

The second test (Armitage & Hills[1982]) carries this idea a stage further by including the 01 and 10 responses and testing for trend in the 2x3 contingency table:

response:	00	01 or 10	11
seq. 1	n_{00}	$n_{01} + n_{10}$	n_{11}
seq. 2	n'_{00}	$n'_{01} + n'_{10}$	n'_{11}

The test used was proposed by Armitage[1955], and is analagous to the "due to linear regression" term in the analysis of variance for a regression. In this case, the

"regression" is of the proportion in sequence 1 of the three types of response on a dummy variable taking values -1, 0, and 1 which characterises the three types of response. Because these variables will not be normally distributed, a χ^2 statistic is defined, rather than using the normal F-test. The basis of this test is very similar to that used by Prescott, but instead of using a randomisation test, a χ^2 equivalent to the usual F-test is used.

6.9 Discussion

Models for cross-over designs with binary data have been considered in detail in this chapter. Most of the models considered have used the logit transformation, with a linear model for the logit of the probability of a positive response for each treatment-period combination. The obvious choice of making this linear model an exact analogue of the model used for continuous responses proves to be inconvenient, with the subject effects being nuisance parameters. However, if subject parameters are omitted from the model, the problem of the dependence between the two observations on the same subjects arises. In some models, this problem has simply been ignored, and no attempt has been made to allow for the dependence. This is clearly not very satisfactory, but the solution adopted by Kenward & Jones is also open to criticism, on the grounds that it is arbitrary.

Chapter 7: Higher order Cross-overs with Binary Data

7.1 Introduction

In chapter 6, possible models for the analysis of the simple two-period cross-over with binary data were reviewed. In most cases, these models can be generalised and applied to more complex cross-over designs with little difficulty. In this chapter, such generalisations will be considered for the cross-over designs considered in chapters 4 & 5, i.e. the "complete" two-period cross-over, with sequences AA, BB, AB & BA; and the three-period designs with sequences ABB & BAA, and with sequences AAA, BBB, ABB & BAA. Baseline observations with binary data are not always sensible, particularly when the binary clasifications are changes in the condition of the patient for example "improved" and "not improved", and will not be considered in detail.

7.2 The Complete Two-period Cross-over

In chapter 4, it was found that, with continuous response variables, the "complete" two-period cross-over, with sequences AA, BB, AB & BA, allowed unbiased estimates of both the treatment and first-order carry-over effects to be made. In addition, use of the two sequences AA & AB only, allowed an unbiased estimate of the treatment effect to be made, even if carry-over was present. It will be shown that these properties also

apply when the observations are binary. Using a model similar to that employed by Gart[1969] for the simple two-period design, the logit of the probability of success for the observation in each period is expressed as a linear function of the overall mean, subject, period, treatment and carry-over effects, in a manner exactly analagous to the model for continuous responses. Because of the inclusion of subject parameters, the probabilities for the two observations for a single subject can be multiplied to give the joint probability of a particular pattern of responses. This gives the following probabilities for subject j in each of the four sequences:

sequence AA $k_{1j}\ p(00) = 1$

$k_{1j}\ p(10) = \exp\{\mu + \omega_{1j} + \pi_1 + \tau_A\}$

$k_{1j}\ p(01) = \exp\{\mu + \omega_{1j} + \pi_2 + \tau_A + \alpha_A\}$

$k_{1j}\ p(11) = \exp\{2\mu + 2\omega_{1j} + 2\tau_A + \alpha_A\}$

where $k_{1j} = (1 + \exp\{\mu + \omega_{1j} + \pi_1 + \tau_A\})(1 + \exp\{\mu + \omega_{1j} + \pi_2 + \tau_A + \alpha_A\})$ ..7.2.1

sequence BB $k_{2j}\ p(00) = 1$

$k_{2j}\ p(10) = \exp\{\mu + \omega_{2j} + \pi_1 + \tau_B\}$

$k_{2j}\ p(01) = \exp\{\mu + \omega_{2j} + \pi_2 + \tau_B + \alpha_B\}$

$k_{2j}\ p(11) = \exp\{2\mu + 2\omega_{2j} + 2\tau_B + \alpha_B\}$

where $k_{2j} = (1 + \exp\{\mu + \omega_{2j} + \pi_1 + \tau_B\})(1 + \exp\{\mu + \omega_{2j} + \pi_2 + \tau_B + \alpha_B\})$ ..7.2.2

sequence AB $k_{3j}\ p(00) = 1$

$k_{3j}\ p(10) = \exp\{\mu + \omega_{3j} + \pi_1 + \tau_A\}$

$k_{3j}\ p(01) = \exp\{\mu + \omega_{3j} + \pi_2 + \tau_B + \alpha_A\}$

$k_{3j}\ p(11) = \exp\{2\mu + 2\omega_{2j} + \alpha_A\}$

where $k_{3j} = (1 + \exp\{\mu + \omega_{3j} + \pi_1 + \tau_A\})(1 + \exp\{\mu + \omega_{3j} + \pi_2 + \tau_B + \alpha_A\})$ ..7.2.3

sequence BA $k_{4j}\ p(00) = 1$

$k_{4j}\ p(10) = \exp\{\mu + \omega_{4j} + \pi_1 + \tau_B\}$

$k_{4j}\ p(01) = \exp\{\mu + \omega_{4j} + \pi_2 + \tau_A + \alpha_B\}$

$k_{4j}\ p(11) = \exp\{2\mu + 2\omega_{4j} + \alpha_B\}$

where $k_{4j} = (1 + \exp\{\mu + \omega_{4j} + \pi_1 + \tau_B\})(1 + \exp\{\mu + \omega_{4j} + \pi_2 + \tau_A + \alpha_B\})$ ..7.2.4

It can be seen that, apart from the normalising constants $k_{1j}, \ldots k_{4j}$, the treatment and carry-over effects only appear in the expressions for the 10 and 01 patterns. Following the procedure adopted by Gart[1969], the subject effects can be eliminated by conditioning on their sufficient statistic, which is the number of successes. The expressions for the probability of success in period one, given there is only one success in the two periods, is as follows for the four sequences:

seq. AA $\exp\{\pi_1\}/(\exp\{\pi_1\} + \exp\{\pi_2 + \alpha_A\})$

seq. BB $\exp\{\pi_1\}/(\exp\{\pi_1\} + \exp\{\pi_2 + \alpha_B\})$

seq. AB $\exp\{\pi_1 + \tau_A\}/(\exp\{\pi_1 + \tau_A\} + \exp\{\pi_2 + \tau_B + \alpha_A\})$

seq. BA $\exp\{\pi_1 + \tau_B\} / (\exp\{\pi_1 + \tau_B\} + \exp\{\pi_2 + \tau_A + \alpha_B\})$

...7.2.5

Clearly, the subject effects have been eliminated, and these conditional probabilities can form the basis of hypothesis tests and estimates. It can be seen that the expressions for sequences AA and BB will be identical if $\alpha_A = \alpha_B$. Thus a test of the equality of these two conditional probabilities is a test of whether carry-over effects can be omitted from the model. The test can be carried out by the usual χ^2 test associated with the 2x2 table:

	success in	
sequence	per 1	per 2
AA	m_{10}	m_{01}
BB	m'_{10}	m'_{01}

where m_{10}, m_{01} are the observed number of subjects in sequence AA giving the response patterns 10 & 01 respectively, and m'_{10}, m'_{01} are the corresponding observed frequencies in sequence BB.

If $\tau_A = \tau_B$, the conditional probabilities for sequences AA & AB, and for sequences BB & BA, will be identical, whether or not carry-over is in the model. Testing the hypothesis of no difference in treatment effects can thus be achieved by considering the 2x2x2

table with classifications Q = sequence pair (AA, AB or BB, BA), S = sequence within pair (parallel or cross-over) and P = period with the success. If there is no treatment effect, there will be no S x P interaction. This can be tested using a log-linear model fitted using a package such as GLIM.

It should be noted that, if only one sequence pair is used, for example AA & AB, a test of the treatment effects is possible via a 2x2 table with the classifications S & P. This test is valid whether or not there is significant carry-over, but a separate test of carry-over is not possible. If there are neither treatment nor carry-over effects, the conditional probabilities for sequences AB & BA will be the same. Hence for the 2x2x2 table described above, there will be no Q x S or Q x P interaction. There should in no circumstances be a three-factor interaction in this 2x2x2 table, and the existance of such an interaction would cast doubt on the form of the model.

By equating the observed relative frequency of success in period 1 given only one success with the theoretical model for this conditional probability, it is possible to obtain estimates of the effects in the model. Writing m_{10}, m_{01} for the observed frequencies of the 10 and 01 patterns in the AA sequence, with m'_{10}, m'_{01}; n_{10}, n_{01}; and n'_{10}, n'_{01} for the corresponding observed frequencies in the BB, AB and BA sequences respectively,

it can be seen that $m_{10}/(m_{10} + m_{01})$ estimates

$\exp(\pi_1)/(\exp(\pi_1) + \exp(\pi_2 + \alpha_A))$

By writing $\pi = \pi_1 = -\pi_2$, and $\alpha = \alpha_A = -\alpha_B$, the expression being estimated can be simplified to $1/(1 + \exp(-2\pi + \alpha))$ so that m_{01}/m_{10} must be an estimate of $\exp(-2\pi + \alpha)$. Similarly, m'_{01}/m'_{10} is an estimate of $\exp(-2\pi - \alpha)$, so that $\ln(m_{01}m'_{10}/m_{10}m'_{10})$ is an estimate of 2α, or $\alpha_A - \alpha_B$. Similarly, by writing $\tau = \tau_A = -\tau_B$, it can be shown that n_{01}/n_{10} estimates $\exp(-2\pi - 2\tau + \alpha)$, and n'_{01}/n'_{10} estimates $\exp(-2\pi + 2\tau - \alpha)$. Thus, $\ln(m_{01}n_{10}/m_{10}n_{01})$ and $\ln(m'_{10}n'_{01}/m'_{01}n'_{10})$ both estimate 2τ, or $\tau_A - \tau_B$. If only sequences AA and AB are used, the first of these two estimators can be used to estimate the treatment difference, but if all four sequences are used, the two estimators can be combined to give $\frac{1}{2}\ln(m_{01}m'_{10}n_{10}n'_{01}/m_{10}m'_{01}n_{01}n'_{10})$ as the estimator of 2τ.

The disadvantage of including subject effects in the model is that, by conditioning on the sufficient statistic to eliminate these nuisance parameters, one neccessarily eliminates the patterns consisting of two successes, or two failures. As with the simple two-period cross-over, this means discounting half of the possible patterns, which will probably lead to a drastic reduction in the number of observations included in the analysis. To avoid this, it would be possible to adapt Prescott's test to the present design. The reasoning behind Prescott's test was that the response patterns 10, 00 or

11, and 01 are progressively less favourable to treatment A, and more favourable to treatment B in the AB sequence, and have the opposite interpretation for the BA sequence. As far as the AA and BB sequences go, a trend in the number of responses to this set of patterns would suggest the operation of carry-over, with any trend being expected to be in the opposite direction for the two sequences. Thus a test of trend for these two sequnces, either using the randomisation test used by Prescott, or the χ^2 test used by Armitage, will give a test of carry-over.

It would also be possible to generalise Kenward & Jones' model to this design. The model for the logit of the probability of success for each period would be similar to that used above, except that subject parameters would not be used. When these are combined to give an expression for the log of the probability of a particular pattern of responses (e.g. 00, 01 etc.) a parameter for the overall tendency for a success to occur is introduced which is modified by dependency parameters. Since the probabilities for the four possible patterns within each sequence may be considered as forming a multinomial distribution, only two dependency parameters can be allowed, in addition to parameters for period, treatment, carry-over and tendency of success, in order to avoid over-parameterisation. The two dependency parameters could be defined in several different ways,

but there seems no reason to change from Kenward & Jones' original parameterisation of having one parameter (σ) which is positive if the responses for the two periods agree, and the other (ϕ) which represents a difference in dependency between the two sequences in a complementary set (AA & BB, and AB & BA). With this parameterisation, the expressions for the log of the probability of the occurence of each pattern within each sequence are:

seq 1 (AA)

00 $\mu_1 + \sigma + \phi$

01 $\mu_1 + \psi + \pi_2 + \tau_A + \alpha_A - \sigma - \phi$

10 $\mu_1 + \psi + \pi_1 + \tau_A - \sigma - \phi$

11 $\mu_1 + 2\psi + \alpha_A + \sigma + \phi$ ..7.2.6

seq 2 (BB)

00 $\mu_2 + \sigma - \phi$

01 $\mu_2 + \psi + \pi_2 + \tau_B + \alpha_B - \sigma + \phi$

10 $\mu_2 + \psi + \pi_1 + \tau_B - \sigma + \phi$

11 $\mu_2 + 2\psi + \alpha_B + \sigma - \phi$ ..7.2.7

seq 3 (AB)

00 $\mu_3 + \sigma + \phi$

01 $\mu_3 + \psi + \pi_2 + \tau_B + \alpha_A - \sigma - \phi$

10 $\mu_3 + \psi + \pi_1 + \tau_A - \sigma - \phi$

11 $\mu_3 + 2\psi + \alpha_A + \sigma + \phi$ ..7.2.8

seq 4 (BA)

00 $\mu_4 + \sigma - \phi$

01 $\mu_4 + \psi + \pi_2 + \tau_A + \alpha_B - \sigma + \phi$

10 $\mu_4 + \psi + \pi_1 + \tau_B - \sigma + \phi$

11 $\mu_4 + 2\psi + \alpha_B + \sigma - \phi$ ..7.2.9

$\mu_1 \ldots \mu_4$ are normalising constants.

A log-linear modelling package can then be used to fit a hierarchy of models containing all or some of these parameters to give tests of the various effects

7.3 Three-period Designs

It has been seen that the main disadvantage of the Gart methodology, of including subject parameters in the model and conditioning on the sufficient statistic to eliminate them, is that the patterns consisting of all successes or all failures cannot be used to obtain information about treatment or carry-over effects. For two period designs, this means that only half the possible patterns are used, but for three-period designs the effect is less severe, as the 000 and 111 patterns which cannot be used only account for a quarter of the possible patterns. Designs with more treatment periods would be even less badly affected, but this advantage might well be negated by the extra expense and the increased probability of drop-outs with a long trial.

7.3.1 The Two-sequence Design

It is a simple matter to write down a Gart-type model for the three period design with sequences ABB and BAA, as the model for the logit of the probability of success is again an exact analogue of the model for continuous responses given in 5.2.1. If we condition on the number of successes, the sufficient statistic for the subject parameters, there are three patterns giving only one success, and three giving two successes with one giving no successes and one giving three successes. In the patterns with only one success, or only one failure, it is natural to characterise the response according to which treatment gives the untypical response, although for the same pattern this is different in the two sequences. An equivalent alternative is to assign a score to each pattern, which characterises the degree to which the pattern supports the hypothesis that treatment A is more effective than treatment B. From the analysis of the continuous case, it would seem that a score formed by taking 2*(response for period 1) - (response for period 2) - (response for period 3) might be useful, where the response is 1 for "success" and 0 for "failure". This gives the ordering given in table 4, which seems logical. It is clear that patterns 100 and 011 are the most extreme, with one treatment giving success, and the other failure. It is also clear that patterns 010 and 001 must favour the treatment given in the second and third

periods, although not as strongly as the 011 pattern. More contentious perhaps, are the 110 and 101 patterns, where each treatment scores one success. In these cases there is some evidence of ineffectiveness in the second treatment, but not for the first, so that it is reasonable to regard the patterns as slightly favouring the first treatment.

Table 4 Implication of patterns in the three-period design

		implication in sequence	
score	patterns	ABB	BAA
2	100	most favours A	most favours B
1	110 101	favours A	favours B
0	000 111	neutral	
-1	010 001	favours B	favours A
-2	011	most favours B	most favours A

It is clear that the patterns 110 & 101 give the same evidence for the comparative efficacy of the two treatments, as do the patterns 010 & 001, so that the occurence of one of the patterns rather than the other might be due to carry-over effects. In order to investigate this, conditioning on the score (+1 or -1) will be necessary rather than on the number of successes. Defining a type R pattern as one in which their is a success in the second period (i.e. 110 or 010), and type

S as one in which their is a success in the third period (i.e. 101 or 001), and using $\tau = \tau_A = -\tau_B$; $\alpha = \alpha_A = -\alpha_B$, the relevant conditional probabilities are:

probability of type R given score =1

for seq 1 (ABB): $\exp(\pi_2 + \alpha)/(\exp(\pi_2 + \alpha) + \exp(\pi_3 - \alpha))$

for seq 2 (BAA): $\exp(\pi_2 - \alpha)/(\exp(\pi_2 - \alpha) + \exp(\pi_3 + \alpha))$

probability of type R given score = -1

for seq 1 : $\exp(\pi_2 + \alpha)/(\exp(\pi_2 + \alpha) + \exp(\pi_3 - \alpha))$

for seq 2 : $\exp(\pi_2 - \alpha)/(\exp(\pi_2 - \alpha) + \exp(\pi_3 + \alpha))$

The two sets of conditional probabilities are identical, a fact that can be used to give a test of the assumptions underlying the model. Defining n_{010}, n_{001} etc. as the observed number of subjects giving patterns 010 , 001 etc. in sequence 1 and n'_{010}, n'_{001} as the corresponding observed frequencies for sequence 2, the test relates to the 2x2x2 contingency table:

		sequence				sequence	
score = -1		1	2	score = +1		1	2
pattern	R	n_{010}	n'_{010}	pattern	R	n_{110}	n'_{110}
type	S	n_{001}	n'_{001}	type	S	n_{101}	n'_{101}

The test of the model is a test of no three-factor or pattern x score interaction, which may be carried out using a log-linear modelling package such as GLIM.

It is clear that all four of the conditional probabilities above will be equal if there is no carry-over effect. This would imply that there is no interaction between patterns and sequences in the contingency table, giving a test for the absence of carry-over. It is also possible to obtain an estimate of the carry-over effect α by equating the observed relative frequencies with the expressions for the conditional probabilities derived from the model. The two relative frequencies relating to sequence 1 both estimate $L = \exp\{\pi_2 + \alpha\}/(\exp\{\pi_2 + \alpha\} + \exp\{\pi_3 - \alpha\})$, while the two conditional probabilities relating to sequence 2 both estimate $M = \exp\{\pi_2 - \alpha\}/(\exp\{\pi_2 - \alpha\} + \exp\{\pi_3 + \alpha\})$. It is natural to combine the two estimators of each of the above expressions in some way. One way would be to weight the two estimators according to the size of their variances, but as the variances contain the parameter we wish to estimate, this is not possible, and a more ad hoc approach must be used. The most reasonable approach is to add the numerators and denominators of the estimators. This gives

$(n_{010} + n_{110})/(n_{010} + n_{001} + n_{110} + n_{101})$ as the estimator of L, and $(n'_{010} + n'_{110})/(n'_{010} + n'_{001} + n'_{110} + n'_{101})$ as the

estimator of M. Since $\ln\{L/(1-L)\}\ln\{M/(1-M)\} = 4\alpha$ the estimator of α is

$$\hat{\alpha} = \frac{1}{4} \ln \left\{ \frac{(n_{010} + n_{110})(n'_{001} + n'_{101})}{(n_{001} + n_{101})(n'_{010} + n'_{110})} \right\} \qquad \ldots 7.3.2$$

In order to provide a test of the treatment effect, it is necessary to condition on the number of successes, rather than the score. This gives the following conditional probabilities:

probability that success is in period 1 given 1 success

$$\text{seq 1} : \frac{\exp\{\pi_1 + \tau\}}{\exp\{\pi_1 + \tau\} + \exp\{\pi_2 - \tau + \alpha\} + \exp\{\pi_3 - \tau - \alpha\}}$$

$$\text{seq 2} : \frac{\exp\{\pi_1 - \tau\}}{\exp\{\pi_1 - \tau\} + \exp\{\pi_2 + \tau - \alpha\} + \exp\{\pi_3 + \tau + \alpha\}}$$

probability that failure is in period 1 given 1 failure

seq 1 :

$$\frac{\exp\{\pi_2 + \pi_3 - 2\tau\}}{\exp\{\pi_2 + \pi_3 - 2\tau\} + \exp\{\pi_1 + \pi_3 - \alpha\} + \exp\{\pi_1 + \pi_2 + \alpha\}}$$

seq 2 :

$$\frac{\exp\{\pi_2 + \pi_3 + 2\tau\}}{\exp\{\pi_2 + \pi_3 + 2\tau\} + \exp\{\pi_1 + \pi_3 + \alpha\} + \exp\{\pi_1 + \pi_2 - \alpha\}}$$

Equating the observed relative frequencies which estimate these conditional probabilities with the above expressions allows algebraic simplification, yielding the following four equations:

$$\frac{n_{100}}{n_{010} + n_{001}} = \frac{\exp\{\pi_1 + \tau\}}{\exp\{\pi_2 - \tau + \alpha\} + \exp\{\pi_3 - \tau - \alpha\}} \qquad \ldots 7.3.3$$

$$\frac{n'_{100}}{n'_{010} + n'_{001}} = \frac{\exp\{\pi_1 - \tau\}}{\exp\{\pi_2 + \tau - \alpha\} + \exp\{\pi_3 + \tau + \alpha\}} \qquad \ldots 7.3.4$$

$$\frac{n_{011}}{n_{101} + n_{110}} = \frac{\exp\{\pi_2 + \pi_3 - 2\tau\}}{\exp\{\pi_1 + \pi_3 - \alpha\} + \exp\{\pi_1 + \pi_2 + \alpha\}} \quad ..7.3.5$$

$$\frac{n'_{011}}{n'_{101} + n'_{110}} = \frac{\exp\{\pi_2 + \pi_3 + 2\tau\}}{\exp\{\pi_1 + \pi_3 + \alpha\} + \exp\{\pi_1 + \pi_2 - \alpha\}} \quad ..7.3.6$$

Dividing equation 7.3.3 by equation 7.3.4 and removing common factors of $\exp\{\tau\}$ and $\exp\{-\tau\}$ from the top and the bottom of the quotient on the R.H.S. of the resulting equation gives:

$$\frac{n_{100}(n'_{010} + n'_{001})}{n'_{100}(n_{010} + n_{001})} = \exp\{4\tau\}\frac{(\exp\{\pi_2 - \alpha\} + \exp\{\pi_3 + \alpha\})}{(\exp\{\pi_2 + \alpha\} + \exp\{\pi_3 - \alpha\})} \quad ..7.3.7$$

Similarly, dividing equation 7.3.5 by equation 7.3.6, removing common factors of $\exp\{-2\tau\}$ and $\exp\{2\tau\}$, and cancelling common factors of $\exp\{\pi_1\}$ from the top and bottom of the quotient on the R.H.S. of the resulting equation gives:

$$\frac{n_{011}(n'_{101} + n'_{110})}{n'_{011}(n_{101} + n_{110})} = \exp\{-4\tau\}\frac{(\exp\{\pi_3 + \alpha\} + \exp\{\pi_2 - \alpha\})}{(\exp\{\pi_3 - \alpha\} + \exp\{\pi_2 + \alpha\})} \quad ..7.3.8$$

Apart from the initial multipliers of $\exp\{4\tau\}$ and $\exp\{-4\tau\}$ the right-hand sides of the above two equations are identical, so that, if $\tau = 0$, the left-hand sides of the two equations should be equal. Testing the equality

of these expressions is thus a test of the treatment effect. The test can be carried out by testing for three-way interaction in the 2x2x2 table

1 success		seq 1	seq 2
in first	no	$n_{010} + n_{001}$	$n'_{010} + n'_{001}$
period?	yes	n_{100}	n'_{100}
1 failure		seq 1	seq 2
in first	no	$n_{101} + n_{110}$	$n'_{101} + n'_{110}$
period?	yes	n_{011}	n'_{011}

It should be noted that, if there is no carry-over, so that $\alpha = 0$, the test of period by sequence interaction in the above table is also a test of $\tau = 0$.

An estimate of τ may be obtained by dividing 7.3.7 by 7.3.8, as the R.H.S. of the resulting quotient is $\exp(8\tau)$. Hence the resulting estimate of τ is:

$$\hat{\tau} = \frac{1}{8} \ln \frac{n_{100}(n'_{010} + n'_{001})n'_{011}(n_{101} + n_{110})}{n'_{100}(n_{010} + n_{001})n_{011}(n'_{101} + n'_{110})} \qquad ..7.3.9$$

Tests and estimates of treatment and first-order carry-over effects are thus possible for this design using a Gart-type model containing individual subject effects. The tests are based on contingency tables and may be conveniently carried out using a log-linear

modelling package such as GLIM. Further details and a worked example are given in Morrey[1989].

Jones & Kenward[1987] have extended their model to the three-period design containing six sequences consisting of all possible orders of three treatments. The model consists of a conventional logit model with parameters for period treatment and first-order carry-over effects, plus a parameter for the overall tendency for the result to be a success, and three dependency parameters. The dependency parameters are labelled σ_{12}, σ_{13}, and σ_{23}, and each is positive in the model if the response for the two periods denoted by the subscipts is the same (i.e. both 0 or both 1). It would be a simple matter to apply such a model to the design under consideration, by removing the subject parameters from the previous model, and introducing the parameter representing the overall tendency of success, and the three dependency parameters. Analysis is by fitting the model via a log-linear modelling package, as before. The advantage of the model over the Gart-type approach is that the responses for the 000 and 111 patterns are still used in the analysis, but the disadvantage of the artificiality of the dependency parameters remains. In both approaches, the introduction of second-order carry-over would cause problems, as with the case of continuous observations.

7.3.2 The Four-sequence Design

In a similar way to the above, logit models for the design with four sequences AAA, BBB, ABB and BAA can be constructed. Once again, the logit of the probability of success for a particular subject in a particular period will be an exact analogue of the linear model used when the design was considered with continuous response variables. However, the score defined with the two-sequence design does not give a sensible ordering of the possible response patterns for the AAA and BBB sequences. For these, three successes is the response pattern most favourable to the treatment being applied, while no successes is the least favourable response. If we condition on the number of successes as before, in order to remove the nuisance parameters representing subject effects, we will effectively prevent any information about the treatment effect being recovered. The relative frequencies of the different patterns giving one, or two, successes tell us nothing about the effectiveness of the treatment in an AAA or BBB sequence, but they may give information about carry-over. The same phenomenon occurs with continuous responses, where, by taking a within subject comparison to remove the subject effect, we necessarily also remove the treatment effect.

If the conditional probabilities "for success in period 1 given only one success", and "failure in period 1 given only one failure" are considered, and the

estimates of these conditional probabilities from the observed frequencies are equated with the expressions for the conditional probabilities derived from the model as before, estimators of the carry-over effect can be obtained. Writing m_{ijk}, and m'_{ijk} for the number of subject showing response i in period 1, response j in period 2 and response k in period 3 (i,j,k = 0 or 1), for sequence AAA and sequence BBB repectively, we obtain:

for sequence AAA

$$\frac{m_{100}}{m_{010} + m_{001}} = \frac{\exp(\pi_1)}{\exp(\alpha)[\exp(\pi_2 + \exp(\pi_3)]} \qquad ..7.3.2.1$$

$$\frac{m_{011}}{m_{110} + m_{101}} = \frac{\exp(\alpha)\exp(\pi_2+\pi_3)}{\exp(\pi_1+\pi_2) + \exp(\pi_1+\pi_3)} \qquad ..7.3.2.2$$

for sequence BBB

$$\frac{m'_{100}}{m'_{010} + m'_{001}} = \frac{\exp(\alpha)\exp(\pi_1)}{\exp(\pi_2) + \exp(\pi_3)} \qquad ..7.3.2.3$$

$$\frac{m'_{011}}{m'_{110} + m'_{101}} = \frac{\exp(\pi_2+\pi_3)}{\exp(\alpha)[\exp(\pi_1+\pi_2) + \exp(\pi_1+\pi_3)]} \quad .7.3.2.4$$

Clearly, dividing 7.3.2.3 by 7.3.2.1 gives an estimator of exp(2α), while another estimator of exp(2α) is provided by dividing 7.3.2.2 by 7.3.2.4. These estimators could be combined by adding their numerators and denominators as before, but it is less clear how they may be combined with the estimator of exp(4α) derived

from the sequences ABB and BAA, which yielded the estimate of α given in 7.3.2. One option would be to multiply the two new estimators to obtain an expression whose expectation will be $\exp(4\alpha)$, and then combine this with the previous estimator of $\exp(4\alpha)$ by adding the numerators and denominators. This gives the rather unwieldly estimator given below:

$$\hat{\alpha} = \frac{1}{4} \ln \frac{a + b}{c + d}$$

where $a = (n_{010} + n_{110})(n'_{001} + n'_{101})$

$b = m_{011}(m_{010} + m_{001})(m'_{110} + m'_{101})m'_{100}$

$c = (n_{001} + n_{101})(n'_{010} + n'_{110})$

$d = m'_{011}(m'_{010} + m'_{001})(m_{110} + m_{101})m_{100}$ ..7.3.2.5

Although the practice of conditioning on the number of successes has allowed us to obtain an improved estimator of carry-over, no information about the treatment effect can be obtained. The benefit of this is somewhat dubious, and it may not be worthwhile running sequences AAA and BBB merely to improve the estimate of carry-over. Alternatively, conditioning on the number of successes could be abandoned for these sequences and the number of successes used to provide information about the relative efficacy of the two treatments. Intuitively, the number of subjects having 3, 2, 1 or 0 successes in the two sequences should show trends with opposite slopes if the

treatment effect is non-zero. This could be tested using a test for trend in the appropriate 2 x 4 contingency table.

7.4 Discussion

It has been seen that Gart's basic method of conditioning on the number of successes to remove subject effects can be applied to higher-order cross-over designs with success, allowing tests and estimators to be derived. The method does have a distinct drawback when sequences of a single treatment (ie AA or AAA) are used, because it does not allow any information about the treatment effect to be obtained from these sequences. It is doubtful whether such sequences are useful when the response is binary.

Chapter 8: Power of Tests

8.1 Introduction

An important characteristic of any significance test is its power, which is defined as the probability that the test rejects the null hypothesis. This probability will be a function of the parameters which are the subject of the test, usually increasing as their value diverges from that specified in the null hypothesis. Thus, in a test of equality of the treatment effects τ_A and τ_B, the null hypothesis will be that $\tau_A - \tau_B = 0$, and an unbiased estimator of $\tau_A - \tau_B$ will be used as the basis of a test of this hypothesis. If it is assumed that the true difference $\tau_A - \tau_B = \Delta_\tau$, and the unbiased estimator of this is $\hat{\Delta}$, with variance σ^2_Δ, and the data is normally distributed, then $\hat{\Delta}$ will also have a normal distribution, provided the estimator is a linear function of the observations. Thus, if z_α is the value of the standard normal variate Z such that $p(-z_\alpha < Z < z_\alpha) = 1 - \alpha$, the test will be such that the null hypothesis will not be rejected in a test at the α% level if:

$$-z_\alpha < \frac{\hat{\Delta}}{\sigma_\Delta} < z_\alpha$$

and hence if:

$$-z_\alpha - \frac{\Delta_\tau}{\sigma_\Delta} < \frac{\hat{\Delta}}{\sigma_\Delta} - \frac{\Delta_\tau}{\sigma_\Delta} < z_\alpha - \frac{\Delta_\tau}{\sigma_\Delta} \qquad 8.1.1$$

Since $(\hat{\Delta} - \Delta_\tau)/\sigma_\Delta$ has a standard normal distribution, the probability that the null hypothesis is rejected will be:

$$1 - \{G(z_\alpha - \Delta_T/\sigma_\Delta) - G(-z_\alpha - \Delta_T/\sigma_\Delta)\} \quad ..8.1.2$$

where G(z) is the probability that a standard normal variate takes a value less than z.

The above applies if there is a single test involved, but in previous chapters it has been shown that testing for a treatment difference with a cross-over design often involves a pre-test to determine whether there is significant carry-over. The exact form of the test for treatment difference depends on the result of the pre-test for carry-over. In general, the situation can be summarised as follows:

Stage 1 : perform pre-test (test P)

Stage 2 : perform main test

If test P is significant use test S

If test P is non-significant use test R

Thus the probability that the test procedure indicates a significant treatment difference is: p[(test P sig & test S sig) or (test P non-sig & test R sig)]. Since test P being significant and test P being non-significant are mutually exclusive events, this may be written:

p[test P sig] * p[test S sig | test P sig] +

p[test P non-sig] * p[test R sig | test P non-sig]

...8.1.3

It will be assumed that the estimators used as the basis for tests P, R, & S are all normally distributed. In order to calculate the power for a given treatment difference Δ_T, it will be necessary to determine whether the tests P, R, & S are independent. If a main test is not independent of the pre-test, the joint probability distribution of the two estimators for the tests will have to be considered, and this will have a bivariate normal distribution. The power of the test procedure will depend on:

(1) the true size of the parameters which are the subject of the pre-test

(2) the significance levels of the pre-test and the main tests

(3) the number of subjects in the experiment

(4) the true size of the difference in treatment effects

It is possible for the situation to become still more complex, if two pre-tests may be required, for example for significant first- and second-order carry-over. This could lead to a sequence of three tests (two pre-tests and the main test), and, if all three tests were related, their joint distribution would form a tri-variate normal distribution.

This chapter will consider the power calculations for the four cross-over designs considered previously, and, for comparison, the parallel and simple two-period

cross-over designs. It will be assumed through-out that the data are normally distributed.

8.2 Calculation of Multivariate Normal Probabilities

As has been pointed out above, if the estimators involved in the pre-test and main test are correlated, it will be necessary to calculate bivariate normal probabilities in order to calculate the power. Further, if there are two pre-tests before the main test, and all three estimators are related, calculation of trivariate normal probabilities will be involved. In this study, a NAG routine was used to calculate univariate normal probabilities, with the t-function described by Owen[1956] being used to facilitate calculation of bivariate normal probabilities, while the S-function described by Steck[1958] was used for calculation of trivariate normal probabilities. Further details of these functions are given in appendix A, and listings of the computer programmes used for power calculations are contained in appendices B & C.

8.3 The Parallel Design

For the purpose of comparison, it is useful to consider the power of the parallel design, in which N subjects are randomly allocated to receive either treatment A or treatment B. For convenience, it will be assumed that N/2 subjects are allocated to each

treatment. In these circumstances, the unbiased estimator of the treatment difference Δ_T will be the difference in the mean response of the two groups, i.e. $\hat{\Delta} = \bar{y}_{1.} - \bar{y}_{2.}$ which has variance $\sigma^2_{\hat{\Delta}} = 4(\sigma^2_S + \sigma^2_e)/N$. Thus the expression for the power given in 8.1.2 contains both the within-subject and between-subject variability. As the between-subject variability is eliminated from many of the tests in cross-over designs, the power for this test cannot be compared with those for cross-over designs unless some assumption is made about the relative sizes of the between-subject and within-subject variation. To avoid making some arbitrary assumption about the relative sizes of these variances, the parallel design with baselines will be considered, allowing the between subject variation to be eliminated. In such a design, the unbiased estimator of the treatment difference is obtained from the contrast between the baseline and first period observations, i.e. $\hat{\Delta} = (\bar{y}_{11.} - \bar{y}_{10.}) - (\bar{y}_{21.} - \bar{y}_{20.})$. This has variance $\sigma^2_{\hat{\Delta}} = 8\sigma^2_e/N$. Thus for a test at the 5% level, 8.1.2 gives the following expression for the power:

$$1 - \{G(1.96 - \frac{\Delta_T\sqrt{N}}{\sigma_e\sqrt{8}}) - G(-1.96 - \frac{\Delta_T\sqrt{N}}{\sigma_e\sqrt{8}})\}$$

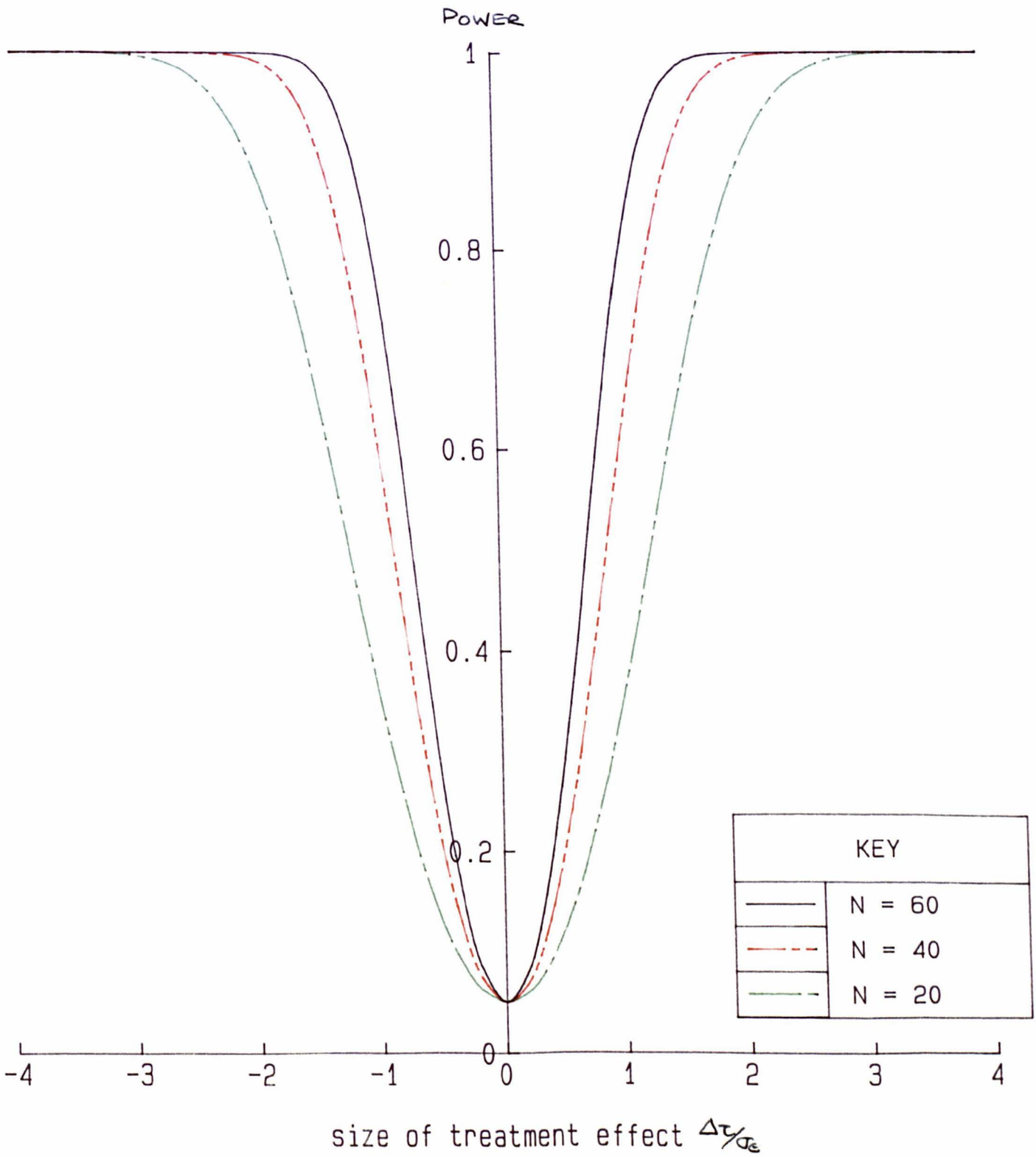
This clearly depends on the sizes of Δ_T, σ_e, and N, of which only N can be controlled by the experimenter. In order to consider the behaviour of the power function for this and other designs, it will be necessary to give a

value for the unknowns Δ_T and σ_e, and the simplest option is to assign specific values to the ratio Δ_T/σ_e and observe the behaviour of the power function. A plot of the power against values of the ratio Δ_T/σ_e between -4 & 4, for N = 20, 40 & 60 is given in figure 1

8.4. The simple two-period cross-over

In order to facilitate comparisons, it is necessary to consider the two-period cross-over with baselines, since, if the presence of carry-over necessitates the deletion of the data from the second period, the test procedure reduces to the parallel with baselines considered above. It will be assumed that N/2 subjects are allocated to each of the two treatment sequences. The required pre-test in this case is for the presence of a significant difference in first-order carry-over, so that the null hypothesis for this pre-test is $\Delta_\alpha = \alpha_A - \alpha_B = 0$. The unbiased estimate of this difference is given by $\hat{\Delta}_\alpha = (2\bar{y}_{20.} - \bar{y}_{21.} - \bar{y}_{22.}) - (2\bar{y}_{10.} - \bar{y}_{11.} - \bar{y}_{12.})$. This has a variance of $24\sigma_e^2/N$. If this test is significant, the test for treatment effect would be the one for the parallel design given above, while if the pre-test is non-significant, the test for treatment would use the contrast between the first and second period observations i.e. the estimator of treatment difference $\hat{\Delta} = \frac{1}{2}\{(\bar{y}_{11.} - \bar{y}_{12.}) - (\bar{y}_{21.} - \bar{y}_{22.})\}$. It should be noted that

Fig 1 : Power of the test for treatment difference for the parallel design with baseline observations



the estimator has expected value $(\tau_A - \tau_B) - \frac{1}{2}(\alpha_A - \alpha_B) = \Delta_\tau - \frac{1}{2}\Delta_\alpha$, so that the estimator is only unbiased for the treatment difference if there is no difference in carry-over.

In order to calculate the power for this suite of tests, it is necessary to determine whether the main tests are independent of the pre-test. If the estimators associated with the three tests are considered as contrasts between the baseline and first and second period observations, the co-efficients for the contrasts are as follows:

	seq 1, period			seq 2, period		
	0	1	2	0	1	2
pre-test(P)	-2	1	1	2	-1	-1
main test(S)	-1	1	0	1	-1	0
main test(R)	0	½	-½	0	-½	½

From this, it is clear that the pre-test and the test used if the pre-test is non-significant (test R) are based on orthogonal contrasts, and so are independent, but that the contrasts for the pre-test and test S are not orthogonal, so that the tests are not independent. It is simple to show that the covariance of the estimators for tests P and S is $12\sigma_e^2/N$, giving a correlation between the estimators of $\sqrt{3}/2$, or 0.866.

Because of the correlation between these two tests, it is necessary to calculate bivariate normal probabilities to evaluate the power function, which will be a function of the true difference in carry-over effects, Δ_α, the true difference in treatment effects, Δ_τ, and the number of subjects, N; as well as the significance levels of the tests. It is common practice to perform the pre-test at a lower significance level than the main test, Grizzle[1965] having recommended the use of a 10% significance level for the pre-test, and a 5% level for the main test. Using these significance levels for the tests, the power for $\Delta_\tau/\sigma_\epsilon = 1$, and for $\Delta_\alpha/\sigma_\epsilon$ between -4 and 4 has been calculated for N = 20, 40 & 60. Tables 5,6 & 7 give the power of the test procedure (i.e. the probability of rejecting the hypothesis that $\Delta_\tau = 0$), the probability that the pre-test for Δ_α is not significant, and the expected value of the estimate of Δ_τ. The results are plotted in figures 2,3 & 4. These results agree with those given by Freeman[1989], although they are presented in a different form.

Table 5 : N = 20 ($n_1 = n_2 = 10$)

$\Delta_\alpha = \alpha_A - \alpha_B$	p(Test P not sig)	Est of Δ_τ	Power
-4.0	0.0224	1.0448	0.3526
-3.8	0.0341	1.0647	0.3527
-3.6	0.0503	1.0906	0.3529

-3.4	0.0723	1.1229	0.3535
-3.2	0.1009	1.1615	0.3551
-3.0	0.1370	1.2055	0.3588
-2.8	0.1811	1.2535	0.3660
-2.6	0.2331	1.3030	0.3793
-2.4	0.2924	1.3509	0.4011
-2.2	0.3580	1.3940	0.4331
-2.0	0.4279	1.4279	0.4761
-1.8	0.5001	1.4501	0.5287
-1.6	0.5721	1.4577	0.5880
-1.4	0.6414	1.4490	0.6502
-1.2	0.7055	1.4233	0.7112
-1.0	0.7626	1.3813	0.7672
-0.8	0.8110	1.3244	0.8149
-0.6	0.8495	1.2549	0.8506
-0.4	0.8774	1.1755	0.8706
-0.2	0.8943	1.0894	0.8706
0.0	0.9000	1.0000	0.8467
0.2	0.8943	0.9106	0.7975
0.4	0.8774	0.8245	0.7260
0.6	0.8495	0.7451	0.6403
0.8	0.8110	0.6756	0.5525
1.0	0.7626	0.6187	0.4748
1.2	0.7055	0.5767	0.4158
1.4	0.6414	0.5510	0.3783
1.6	0.5721	0.5423	0.3599
1.8	0.5001	0.5498	0.3551

2.0	0.4279	0.5721	0.3589
2.2	0.3580	0.6062	0.3672
2.4	0.2924	0.6491	0.3777
2.6	0.2331	0.6970	0.3880
2.8	0.1811	0.7465	0.3962
3.0	0.1370	0.7945	0.4006
3.2	0.1009	0.8385	0.4003
3.4	0.0723	0.8771	0.3957
3.6	0.0503	0.9094	0.3884
3.8	0.0341	0.9353	0.3801
4.0	0.0224	0.9552	0.3722

Table 6 : N = 40 ($n_1 = n_2 = 20$)

$\Delta_\alpha = \alpha_A - \alpha_B$	p(Test P not sig)	Est of Δ_τ	Power
-4.0	0.0002	1.0004	0.6088
-3.8	0.0006	1.0011	0.6088
-3.6	0.0013	1.0024	0.6088
-3.4	0.0030	1.0051	0.6088
-3.2	0.0065	1.0103	0.6088
-3.0	0.0129	1.0194	0.6088
-2.8	0.0244	1.0342	0.6088
-2.6	0.0435	1.0565	0.6088
-2.4	0.0730	1.0876	0.6088
-2.2	0.1160	1.1276	0.6089
-2.0	0.1743	1.1743	0.6094

-1.8	0.2485	1.2237	0.6114
-1.6	0.3369	1.2695	0.6175
-1.4	0.4351	1.3046	0.6320
-1.2	0.5374	1.3224	0.6600
-1.0	0.6366	1.3183	0.7036
-0.8	0.7260	1.2904	0.7593
-0.6	0.8001	1.2400	0.8190
-0.4	0.8551	1.1710	0.8736
-0.2	0.8887	1.0889	0.9167
0.0	0.9000	1.0000	0.9448
0.2	0.8887	0.9111	0.9542
0.4	0.8551	0.8290	0.9393
0.6	0.8001	0.7600	0.8946
0.8	0.7260	0.7096	0.8222
1.0	0.6366	0.6817	0.7378
1.2	0.5374	0.6776	0.6651
1.4	0.4351	0.6954	0.6204
1.6	0.3369	0.7305	0.6037
1.8	0.2485	0.7763	0.6036
2.0	0.1743	0.8257	0.6091
2.2	0.1160	0.8724	0.6148
2.4	0.0730	0.9124	0.6188
2.6	0.0435	0.9435	0.6203
2.8	0.0244	0.9658	0.6193
3.0	0.1293	0.9806	0.6166
3.2	0.0065	0.9897	0.6137
3.4	0.0030	0.9949	0.6114

3.6	0.0013	0.9976	0.6100
3.8	0.0006	0.9989	0.6093
4.0	0.0002	0.9996	0.6090

Table 7 : N = 60 ($n_1 = n_2 = 30$)

$\Delta_\alpha = \alpha_A - \alpha_B$	p(Test P not sig)	Est of Δ_τ	Power
-4.0	0.0000	1.0000	0.7819
-3.8	0.0000	1.0000	0.7819
-3.6	0.0000	1.0000	0.7819
-3.4	0.0001	1.0002	0.7819
-3.2	0.0003	1.0005	0.7819
-3.0	0.0010	1.0015	0.7819
-2.8	0.0027	1.0038	0.7819
-2.6	0.0068	1.0089	0.7819
-2.4	0.0158	1.0189	0.7819
-2.2	0.0333	1.0367	0.7819
-2.0	0.0646	1.0646	0.7819
-1.8	0.1148	1.1033	0.7819
-1.6	0.1881	1.1504	0.7819
-1.4	0.2847	1.1993	0.7822
-1.2	0.4001	1.2401	0.7837
-1.0	0.5247	1.2624	0.7893
-0.8	0.6462	1.2585	0.8043
-0.6	0.7521	1.2256	0.8332
-0.4	0.8329	1.1666	0.8739

-0.2	0.8831	1.0883	0.9167
0.0	0.9000	1.0000	0.9517
0.2	0.8831	0.9117	0.9740
0.4	0.8329	0.8334	0.9822
0.6	0.7521	0.7744	0.9722
0.8	0.6462	0.7415	0.9376
1.0	0.5248	0.7376	0.8806
1.2	0.4001	0.7599	0.8214
1.4	0.2847	0.8007	0.7843
1.6	0.1881	0.8496	0.7738
1.8	0.1148	0.8967	0.7769
2.0	0.0646	0.9354	0.7815
2.2	0.0333	0.9633	0.7841
2.4	0.0158	0.9811	0.7849
2.6	0.0068	0.9911	0.7845
2.8	0.0027	0.9962	0.7835
3.0	0.0010	0.9985	0.7827
3.2	0.0003	0.9995	0.7822
3.4	0.0001	0.9998	0.7820
3.6	0.0000	1.0000	0.7819
3.8	0.0000	1.0000	0.7819
4.0	0.0000	1.0000	0.7819

Fig 2 : Performance of the test procedure for the two-period cross-over. N = 20 ($n_1 = n_2 = 10$)

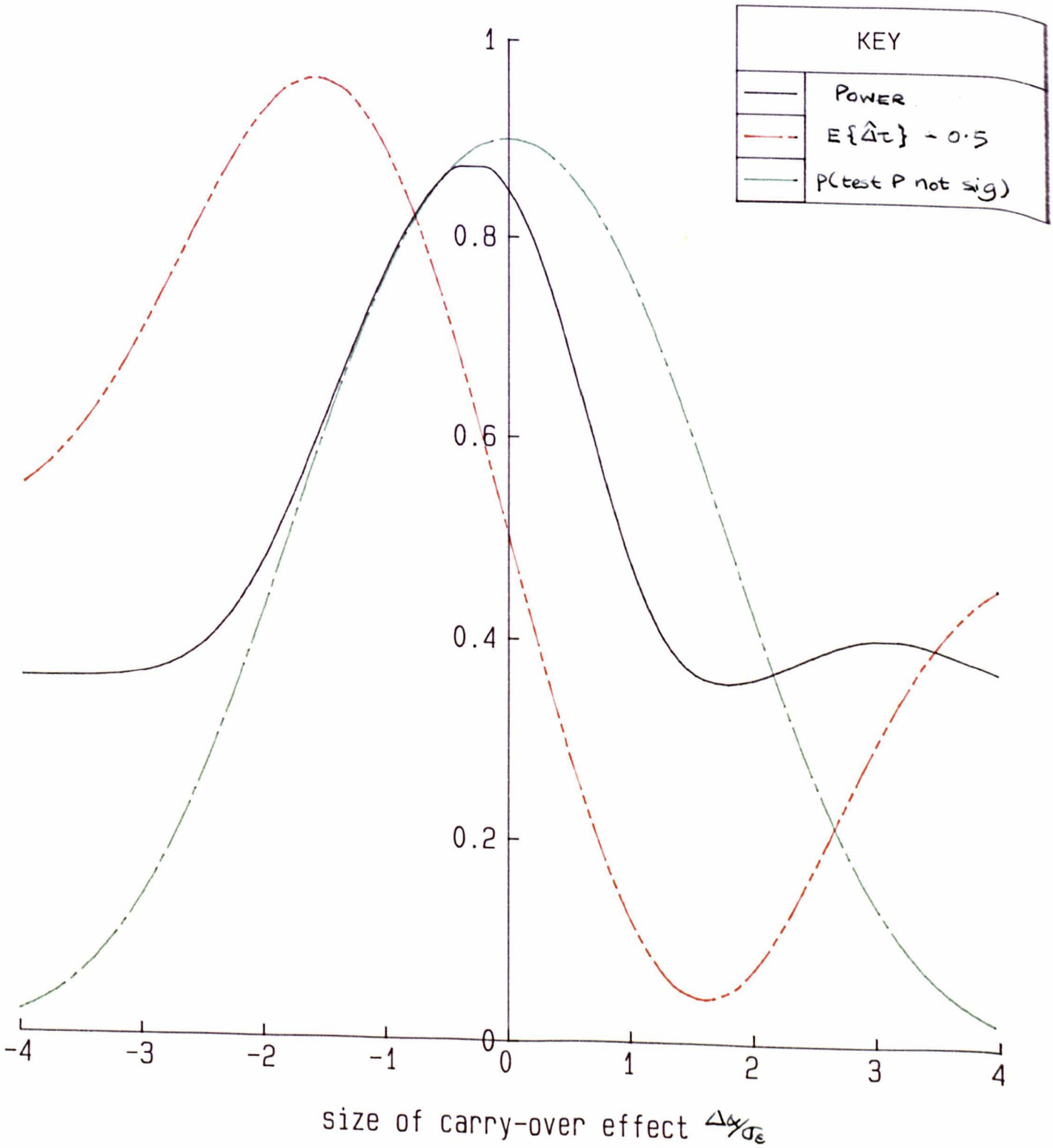


Fig 3 : Performance of the test procedure for the two-period cross-over. N = 40 ($n_1 = n_2 = 20$)

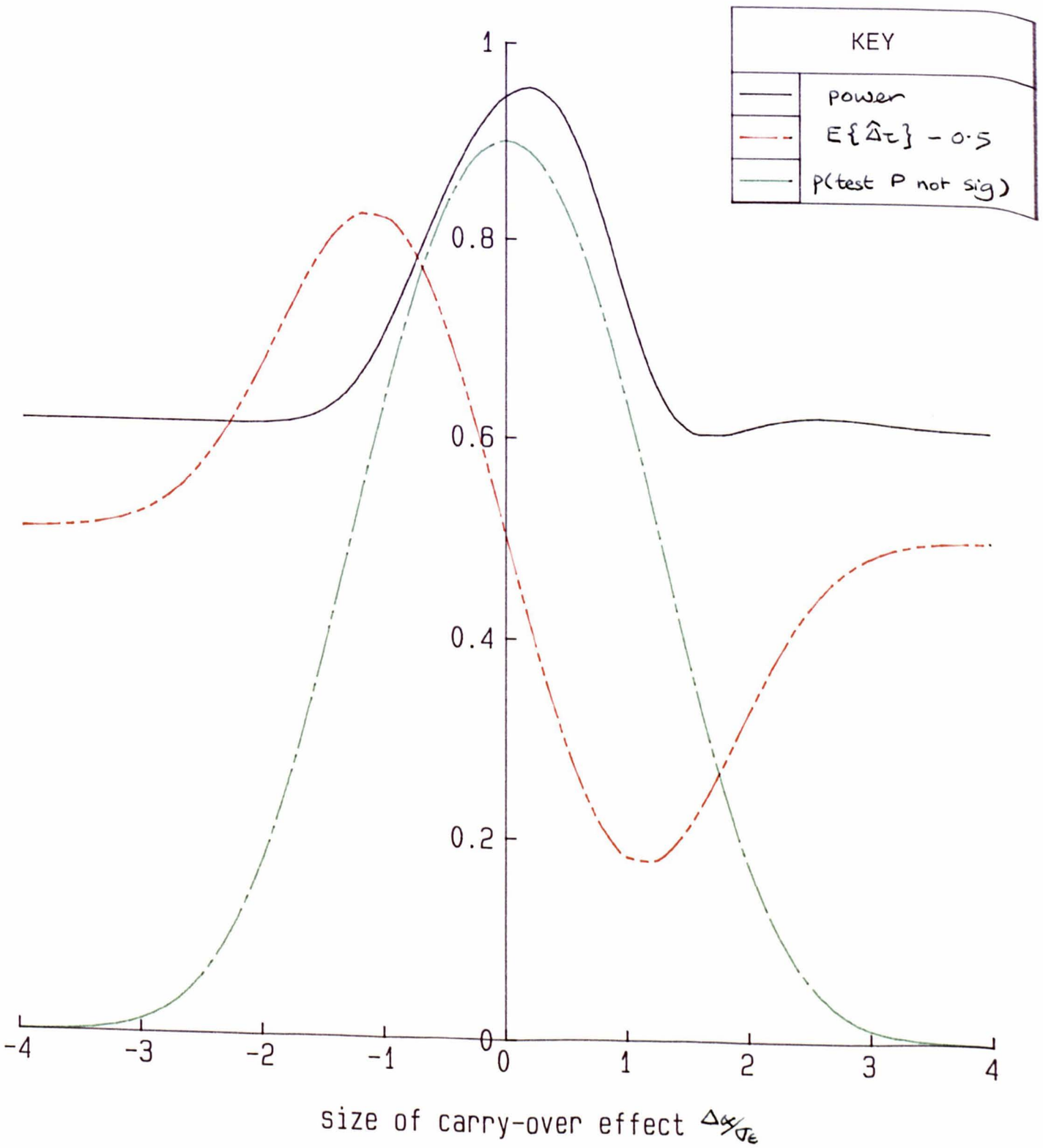
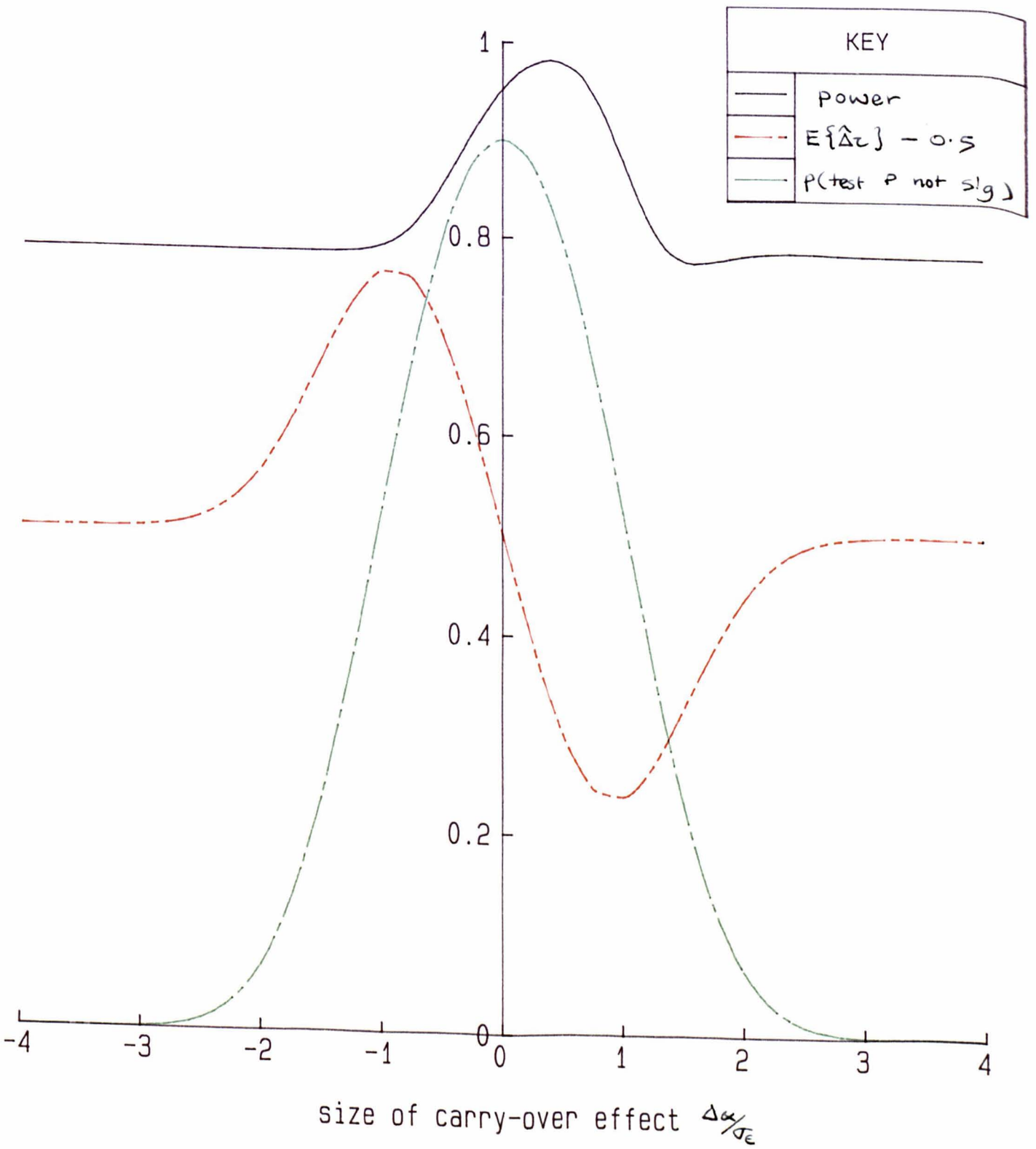


Fig 4 : Performance of the test procedure for the two-period cross-over. N = 60 ($n_1 = n_2 = 30$)



It can be seen that, for large values of $\Delta_\alpha/\sigma_\epsilon$, the power is very similar to that of the parallel design, for which the power to detect a difference of $\Delta_\tau = 1$ is 0.3526 when N = 20, 0.6088 when N = 40 and 0.7819 when N = 60. This is because the pre-test is almost certain to be significant, so that there is a very high chance that the main test will be the one used in the parallel design. Similarly, if $\Delta_\alpha/\sigma_\epsilon$ is close to zero, there is a high chance that the pre-test will be non-significant, and that the more powerful cross-over test will be used, giving a larger value for the power than the parallel design. For moderate values of $\Delta_\alpha/\sigma_\epsilon$, the pre-test may or may not be significant, so that either of the main tests might be used. The advantage of using the cross-over test is less clear-cut because the estimator on which it is based is biased. As Freeman[1989] pointed out, there is a sense in which increasing the sample size does not help, because the size of the carry-over difference Δ_α which causes problems merely decreases, rather than the problem disappearing. It can be seen from tables 5..6 that the largest bias in the estimate of Δ_τ occurs at approximately the value of Δ_α which gives an equal chance of the pre-test being significant. As the sample sizes are increased, this value also decreases.

Lehmacher[1991] has pointed out that the signs of the treatment and carry-over differences are important. Since the expectation of the estimator of the treatment

difference for the cross-over test (test R) is $\Delta_\tau - \frac{1}{2}\Delta_\alpha$, the estimate of the treatment difference is larger than the actual treatment difference if Δ_τ and Δ_α have opposite signs, or if Δ_α is more than four times as large as Δ_τ, if they have the same sign. As far as the power is concerned, this will be beneficial, since the increase in the estimate of the treatment difference makes it more likely that the treatment difference will be detected. On the other hand, if Δ_τ and Δ_α have the same sign, and Δ_α is less than four times the size of Δ_τ, the estimate of the treatment difference will be smaller than the actual treatment difference, shrinking to zero if $\Delta_\alpha = 2\Delta_\tau$. This will reduce the power by making the treatment difference less easy to detect.

It should be noted that this bias only applies to the estimate of the treatment difference used in the cross-over test. If the carry-over difference is large (i.e. four times the size of the treatment difference) the pre-test is likely to be significant, so that the cross-over test is unlikely to be used. However, if the carry-over difference is small, the pre-test is likely to be non-significant, so that the biased estimator of the treatment difference is used. If carry-over was pharmacological in nature, then the sign of the carry-over difference would probably be the same as treatment difference, with the more effective treatment retaining more of its effectiveness in the next period, but in

these circumstances, it is difficult to believe that the size of the carry-over difference could be much larger than the size of the treatment difference. If, on the other hand, carry-over was psychological in nature, then the sign of the carry-over difference could be expected to be different from the sign of the treatment difference because a good treatment might make a rather worse treatment that followed it seem even more inferior, and vice versa. Once again, it is difficult to imagine a case where the carry-over difference would be large compared to the treatment difference. Because pharmacalogical carry-over is relatively easy to avoid, carry-over is more likely to have a psychological origin, or to be some non-specific treatment-period interaction. The previous argument suggests that the carry-over difference is unlikely to be four times as large as the treatment difference, and is more likely to have the opposite sign, so that the bias in the estimate of the treatment effect will probably lead to an over-estimation, but this is unlikely to have much of a beneficial effect on the power to detect a treatment difference.

8.5 The "Complete" Two-period Cross-over

In section 4.3, it was shown that the "complete" two-period cross-over, with four sequences AA, BB, AB, & BA, could give an unbiased estimator of the treatment difference even if the carry-over difference was non-

zero. Thus, if this design is employed, there is no need to perform a pre-test for carry-over before the main test for difference in the treatments, and the power of the test for treatment difference can be calculated very simply, using the formula given in 8.1.1. Following the notation in section 4.3, AA will be denoted sequence 1, BB sequence 2, AB sequence 3 and BA sequence 4; with $\bar{D}_{i.}$ denoting the average difference between the first and second period observations for the subjects in sequence i. If the N subjects are allocated equally to the four sequences, giving N/4 in each sequence, the unbiased estimator of the treatment difference is $\hat{\Delta} = \frac{1}{2}\{\bar{D}_{3.} - \bar{D}_{1.} + \bar{D}_{2.} - \bar{D}_{4.}\}$, which has variance $\sigma^2_{\hat{\Delta}} = 8\sigma^2_e/N$. Since this variance is identical to that for the parallel design with baselines, the power function for this design will be identical to the function graphed in figure 1.

In section 4.3.1 it was noted that the addition of baseline observations to this design could give an unbiased estimator of the treatment difference with a smaller variance. Defining $\bar{G}_{i.}$ to be the average value of the combination $y_{ij0} + y_{ij1} - 2y_{ij2}$ for the subjects in sequence i, the unbiased estimator is $\hat{\Delta} = \frac{1}{4}\{\bar{G}_{3.} - \bar{G}_{1.} + \bar{G}_{2.} - \bar{G}_{4.}\}$ and has variance $\sigma^2_{\hat{\Delta}} = 6\sigma^2_e/N$. Thus, for a test at the 5% significance level, the power will be given by:

$$1 - \{G(1.96 - \frac{\Delta_T\sqrt{N}}{\sigma_e\sqrt{6}}) - G(-1.96 - \frac{\Delta_T\sqrt{N}}{\sigma_e\sqrt{6}})\}$$

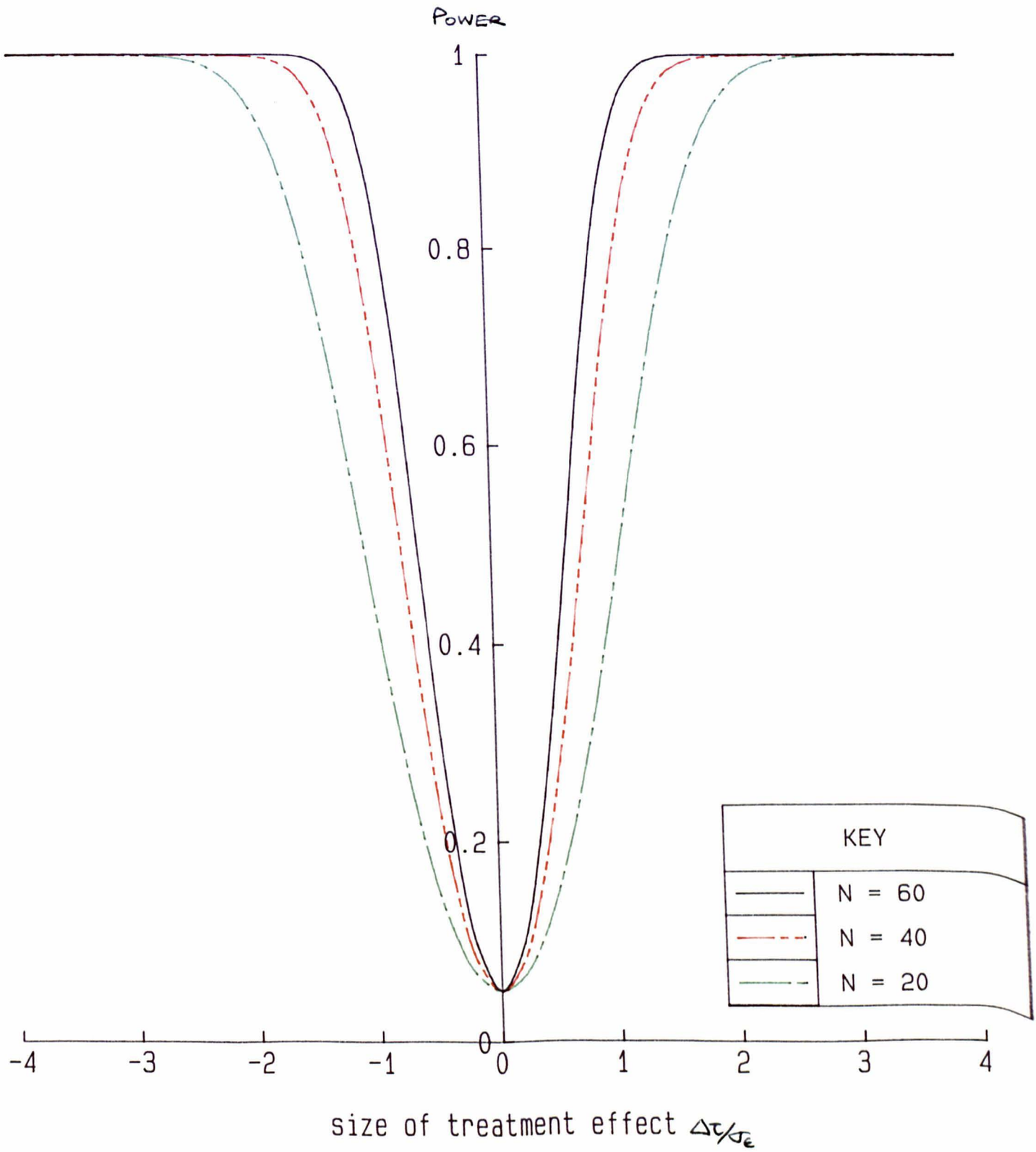
This power function is plotted for values of Δ_T/σ_ϵ between -4 and 4 and for N = 20, 40 & 60 in figure 5.

8.6 The Three-period Designs

Two three-period designs have been considered previously, one having the two sequences ABB and BAA, and the other having four sequences AAA, BBB, ABB and BAA. With both of these designs, the possibility of second-order carry-over effects must be considered. If the form of the test for treatment difference depends on whether or not there is any difference in the second-order carry-over effects, a pre-test for this will be necessary. Another pre-test for first-order carry-over may also be necessary so that it is possible that the test procedure would consist of a set of three tests: two pre-tests and the main test for treatment difference. If all three tests are correlated it is necessary to evaluate a tri-variate normal probability to calculate the chance of a particular sequence of results for these three tests. Hence, when the sequence of three tests lead to a decision that there is a significant difference between the treatments, a tri-variate normal probability is required to calulate the contribution to the overall power.

In view of the problems found with the simpler sequence of tests for the two-period cross-over, it is interesting to examine how this more complex suite of

Fig 5 : Power of the test for treatment difference for the "complete" two-period cross-over



tests performs. The situation was not studied by Freeman[1989] and results for power calculations for this suite of tests have not been published elsewhere.

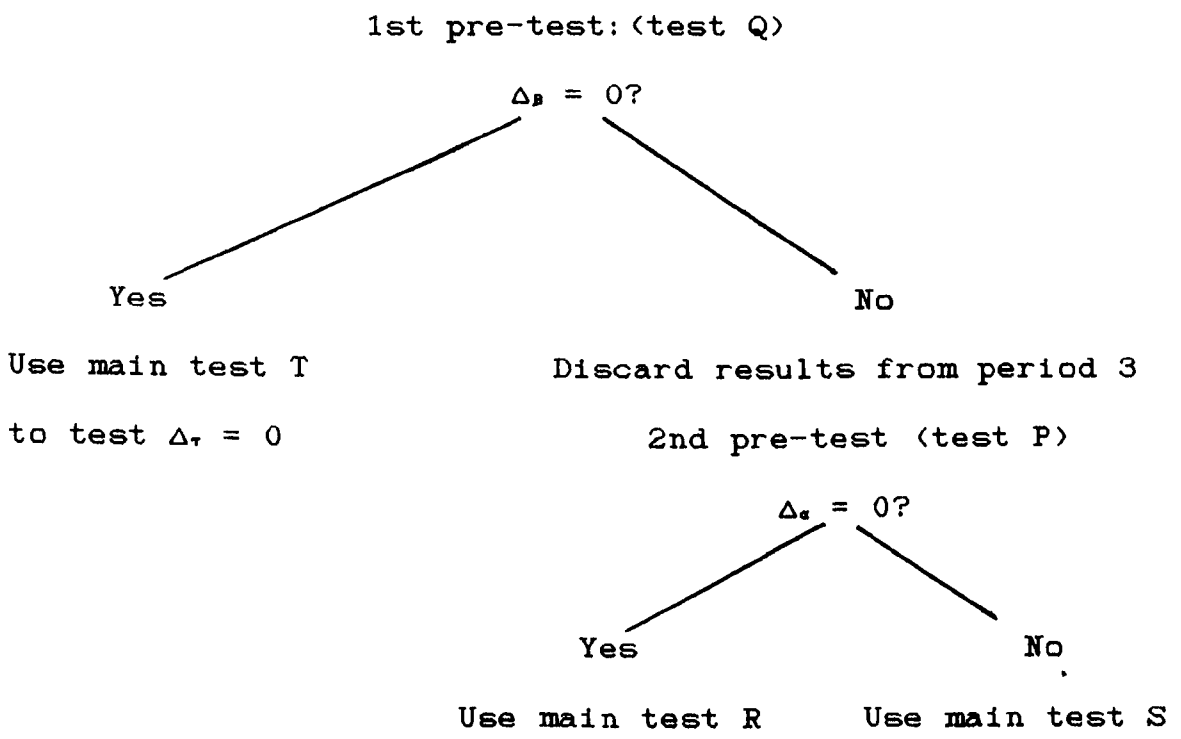
8.6.1 The Two-sequence Design

The analysis of the two-sequence design was considered in chapter 5, where it was shown that the estimator of the treatment difference is only unbiased if there is no difference in the second-order carry-over effects. If there are no baseline observations, it is not possible to find a within-subject comparison which will test for second-order carry-over, and, if a between-subject comparison is made, the variance of the estimator would contain the between-subject variance, σ_s^2. As before, in order to avoid the necessity of some assumption about the relative sizes of the between and withing subject variability, the design with baselines will be considered. It should also be noted that, if there is significant difference in second-order carry-over effects, the results from the third treatment period cannot be used, effectively reducing the design to the two-period cross-over, which has already been considered with baselines.

From section 5.2.1 it can be seen that even with baseline observations the estimator of treatment difference with the smallest variance is based on the contrast between the three treatment periods $H_{ij} = 2y_{ij1} -$

$y_{ij2} - y_{ij3}$. Writing $\bar{H}_{i.}$ for the average of the H_{ij} values in sequence i, the estimator of the treatment difference is $\hat{\Delta} = \frac{1}{4}(\bar{H}_{1.} - \bar{H}_{2.})$. However, this has expected value $\tau_A - \tau_B - \frac{1}{4}(\beta_A - \beta_B)$, or $\Delta_\tau - \frac{1}{4}\Delta_\beta$, where Δ_β is the true difference in second-order carry-over effects. If a test for the equality of second-order carry-over effects is non-significant, the above estimator of the treatment difference is valid, but otherwise, the results for the third treatment period must be discarded, leaving results from a two-period cross-over with baselines. Hence, if the pre-test for second-order carry-over is significant, the situation reverts to that given above for the two-period cross-over, where a pre-test for first-order carry-over is necessary in order to determine the appropriate test for treatment difference. The procedure if summarised in figure 6. The contrasts used in the five tests involved in the test procedure are defined below, together with the corresponding estimator and it's expectation. The leading co-efficient of the variance of the estimator are also given. In order to determine whether the estimators on which the tests are based are independent, they may each be regarded as contrasts between the baseline and three treatment period observations of the two groups. The co-efficients for these contrasts are also tabulated.

Fig 6 : Sequence of tests for the three-period two-sequence design



CONTRASTS

Test	Q	$Q_{ij} = 4y_{ij0} - 2y_{ij1} - y_{ij2} - y_{ij3}$
Test	P	$E_{ij} = 2y_{ij0} - y_{ij1} - y_{ij2}$
Test	T	$H_{ij} = 2y_{ij1} - y_{ij2} - y_{ij3}$
Test	R	$D_{ij} = y_{ij1} - y_{ij2}$
Test	S	$F_{ij} = y_{ij0} - y_{ij1}$

ESTIMATORS

	Estimator	Expectation	Variance Co-eff
Test Q	$\bar{Q}_{2.} - \bar{Q}_{1.}$	Δ_β	88
Test P	$\bar{E}_{2.} - \bar{E}_{1.}$	Δ_α	24
Test T	$\frac{1}{4}(\bar{H}_{1.} - \bar{H}_{2.})$	$\Delta_\tau - \frac{1}{4}\Delta_\beta$	1.5

Test R	$\frac{1}{2}(\bar{D}_{1.} - \bar{D}_{2.})$	$\Delta_\tau - \frac{1}{2}\Delta_\alpha$	2
Test S	$\bar{F}_{2.} - \bar{F}_{1.}$	Δ_τ	8

CONTRAST CO-EFFICIENTS

	seq 1, period				seq 2, period			
	0	1	2	3	0	1	2	3
Test Q	-4	2	1	1	4	-2	-1	-1
Test P	-2	1	1	0	2	-1	-1	0
Test T	0	½	-¼	-¼	0	-½	¼	¼
Test R	0	½	-½	0	0	-½	½	0
Test S	-1	1	0	0	1	-1	0	0

It is clear that only the contrasts for tests P and R are orthogonal, so that all other pairs of estimators are correlated. However, for any given experiment, a maximum of three tests will be used (Q, P & R; or Q, P & S), so that only tri-variate normal probabilities will be needed. The probability that the complete test procedure leads to a decision that the treatments are significantly different is:

p{[test Q sig & {(test P sig & test S sig) or (test P non-sig & test R sig)}] or [test Q non-sig & test T sig]}

Taking into account the mutually exclusive events, this may be written:

p[test Q sig] * {p[test P sig | test Q sig] * p[test S sig | Tests Q & P sig] + p[test P non-sig | test Q sig] * p[test R sig | test Q sig & test P non-sig]} + p[test Q non-sig] * p[test T sig | test Q non-sig]

A FORTRAN program was written to calculate this probability, given values of n_1, n_2, Δ_τ, Δ_α, Δ_β, and the significance level of the five tests. A listing of the program is given in appendix C. As with the two-period design, calculations were made using a 10% significance level for the pre-tests of first or second order carry-over (tests Q & P), and a 5% significance level for tests of the treatment effect (tests T, R & S).

The situation is similar to that for the two-period design in as much as the test for a treatment difference is based on a biased estimator if the pre-test fails to detect a non-zero carry-over effect. It is also the case that the expected value of the estimator of the treatment effect is less than the true value if treatment and carry-over effects have the same sign, unless the carry-over effect is considerably larger than the treatment effect.

In order to avoid a sequence of plots for different values of N, it was decided to consider values of $\Delta\ n/\sigma_\epsilon$, where $n = n_1 = n_2 = N/2$. Following Freeman[1989] values of $\Delta_\tau\ n/\sigma_\epsilon$ between -6 and 6 were used, with $\Delta_\alpha\ n/\sigma_\epsilon$ and

$\Delta_\beta\ n/\sigma_\epsilon$ taking either the value zero or 5 n/σ_ϵ. Figure 7 shows the power of the test proceedure for $\Delta_\tau\ n/\sigma_\epsilon$ between -6 and 6 with $\Delta_\alpha\ n/\sigma_\epsilon = \Delta_\beta\ n/\sigma_\epsilon = 0$; $\Delta_\tau\ n/\sigma_\epsilon = 5$ and $\Delta_\beta\ n/\sigma_\epsilon = 0$; and $\Delta_\alpha\ n/\sigma_\epsilon = \Delta_\beta\ n/\sigma_\epsilon = 5$.

As might be expected, the presence of first-order carry-over has little effect in the absence of second-order carry-over, the design being efficient in dealing with first-order carry-over. The effect of second-order carry-over on the power curves is similar to the effect of first-order carry-over on the simple two-period design, with a marked assymetry in the power function, with a "plateau" in the curve for small negative values of the treatment effect $\Delta_\tau\ n/\sigma_\epsilon$.

8.6.2 The Four Sequence Design

In section 5.3, it was seen that the design with four sequences AAA, BBB, ABB & BAA could give an unbiased estimator for Δ_τ even if Δ_α and Δ_β were not zero. Thus there is no need to perform pre-tests for this design and the power of the test for treatment difference has the simple form for a single test. Because the estimator on which the test has a relatively small variance, equal to $7\sigma_\epsilon^2/2N$ if the four groups are each allocated N/4 subjects, or approximately $2.91\sigma_\epsilon^2/N$ if the ratio of subjects in the ABB and BBA sequences to subjects in the AAA and BBB sequences is 5 : 2 or 8 : 3, the test will be correspondingly more powerful than the tests from the

Fig 7 : Performance of the test procedure for the three-period cross-over.

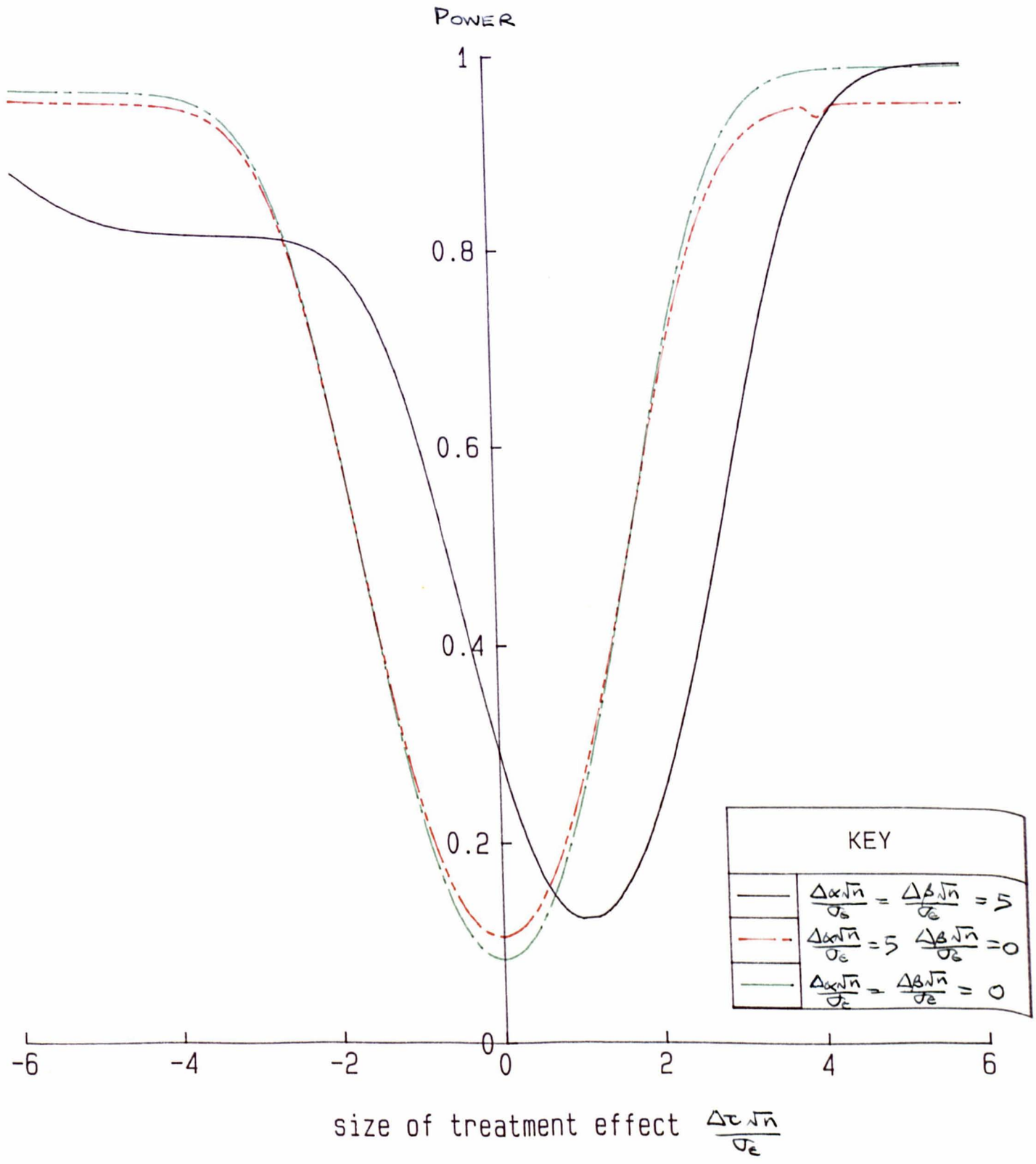
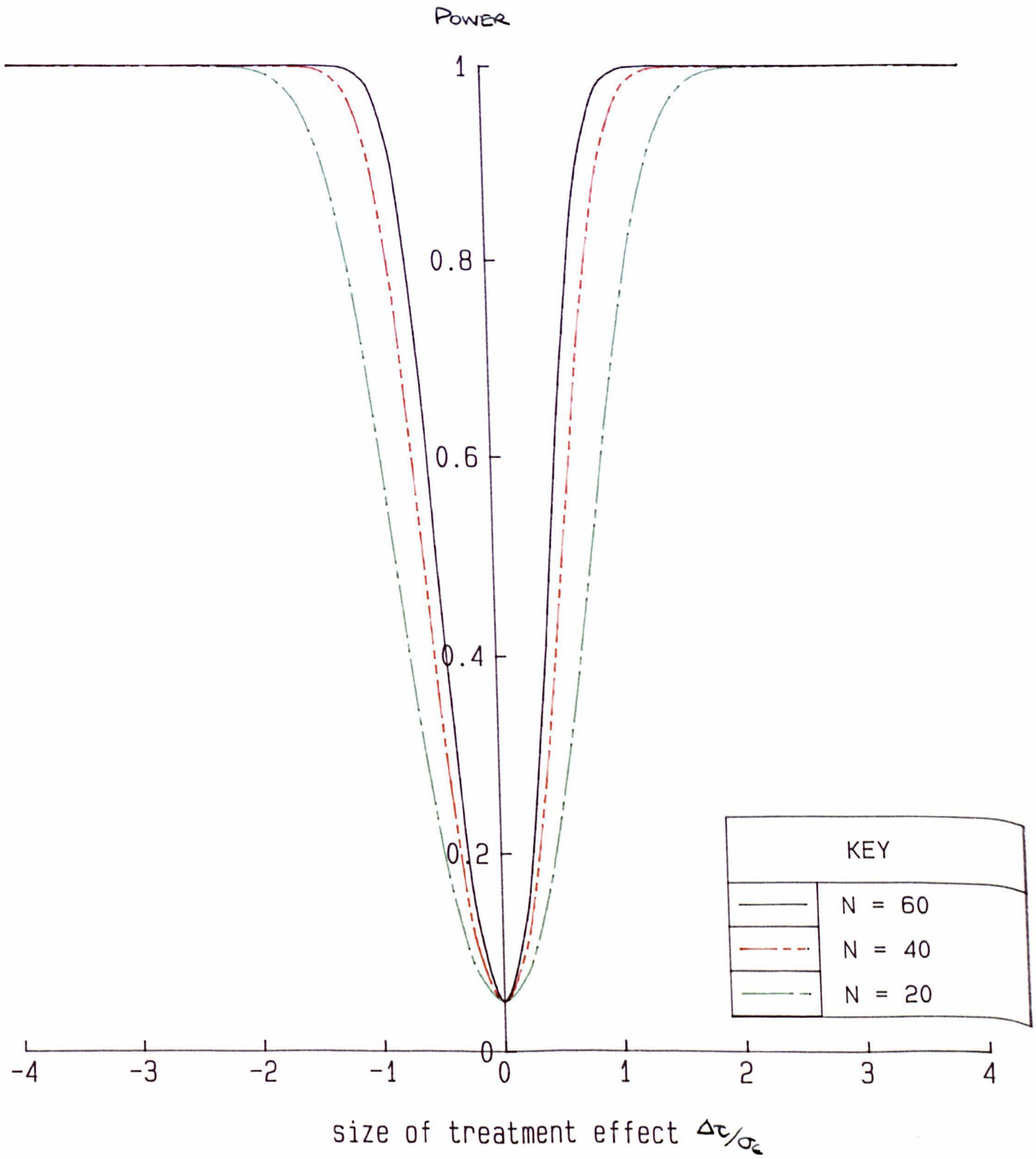


Fig 8 : Power of the test for treatment difference for the four sequence three-period cross-over



parallel design amd the complete two-period cross-over for the same significance level and value of N. A plot of the power function for values of $\Delta_\tau/\sigma_\epsilon$ between -4 and 4, and N = 20, 30 & 40, assuming equal numbers in the four sequences, is given in fig 8.

Discussion

The situation where a sequence of tests is performed, or a pre-test determines the exact form of the main test, is complicated, and the implications for the power of the test proceedure are not intuitively obvious. Indeed, some of the results obtained by Freeman[1989], and confirmed here but presented in a different form, are counter-intuitive. It was often thought that, if the pre-test was powerful enough, no problem would exist, but it has been shown that the problem is merely shifted to occur at smaller values of the carry-over effect. The results for the three-period design presented here show that very similar problems occur with that design, although these are caused by second-order carry-over rather than first-order carry-over. Thus the benefit of the more complex design is in making the problems more distant, and possibly less likely to occur. However, doubt must be cast on the suitability of a test procedure invoving a complex of pre-tests and main tests. One way of avoiding this is to eschew tests altogether and

perform a Bayesian analysis. This option will be explored in the following chapters.

Chapter 9: Bayesian Analysis of Cross-over Designs

9.1 Introduction

In chapters 2 to 5, classical analysis of cross-over designs with continuous response variables was considered. We now consider a Bayesian approach to the analysis of cross-over trials. Such an analysis for the simple two-period cross-over has already been considered by Grieve[1985], Selwyn et al[1981] and Racine et al[1986]. In all of these, the standard linear model is used, as given in 2.2.1, but a prior distribution is postulated for the unknown parameters. By assuming that the data is normally distributed, the likelihood can be written down, and a posterior distribution for the parameters of interest can be calculated using Bayes' Theorem. This chapter will review the results obtained for the simple two-period cross-over, and consider the possibility of extending the method to other cross-over designs. An alternative way of performing the Bayesian analysis using Gibbs sampling will also be considered.

9.2 Results for the Simple Two-period Cross-over

Grieve[1985] has given a Bayesian analysis of the simple two-period cross-over without baseline measurements using a linear model which is identical in form to that given in 2.2.1 above, with the observation for subject j in sequence i for period k being expressed

in terms of an overall mean, μ, an individual subject effect ω, a period effect π, a treatment effect τ, and, for observations in the second period, a carry-over effect α. Making the usual assumptions that $\pi_1 + \pi_2 = \tau_A + \tau_B = \alpha_A + \alpha_B = 0$, the analysis is simplified by defining $\pi = \pi_1 = -\pi_2$, $\tau = \tau_A = -\tau_B$, and $\alpha = \alpha_A = -\alpha_B$. The two observations on subject j in sequence i $(y_{ij1}$ & $y_{ij2})$ are assumed to have a bivariate normal distribution, with $E\{y_{ij1}\}$, $E\{y_{ij2}\}$ given by the linear model, and $var\{y_{ij1}\} = var\{y_{ij2}\} = \sigma_\omega^2 + \sigma_\varepsilon^2$, and $cov\{y_{ij1}, y_{ij2}\} = \sigma_\omega^2$ so that the correlation between the two observations is $\rho = \sigma_\omega^2/(\sigma_\omega^2 + \sigma_\varepsilon^2)$, and $1 - \rho^2 = \sigma_\varepsilon^2\sigma_2^2/(\sigma_\omega^2 + \sigma_\varepsilon^2)^2$, where $\sigma_2^2 = 2\sigma_\omega^2 + \sigma_\varepsilon^2$.

Thus the likelihood for this pair of observations is:

$$1/(2\Pi\sigma_\varepsilon\sigma_2)\exp\{-(\sigma_\omega^2 + \sigma_\varepsilon^2)/2\sigma_\varepsilon^2\sigma_2^2[(y_{ij1} - E\{y_{ij1}\})^2 - 2\sigma_\omega^2/(\sigma_\omega^2 + \sigma_\varepsilon^2)(y_{ij1} - E\{y_{ij1}\})(y_{ij2} - E\{y_{ij2}\}) + (y_{ij2} - E\{y_{ij2}\})^2]\} \quad \ldots 9.2.1$$

Assuming that there are n_i subjects in sequence i, with $N = n_1 + n_2$, $m = N/(n_1n_2)$, the likelihood for the observations on all N subjects can be split into the following components:

$$\hat{\mu} = \tfrac{1}{4}(\Sigma y_{1j1} + \Sigma y_{1j2} + \Sigma y_{2j1} + \Sigma y_{2j2}) \sim N(\mu, m\sigma_2^2/8)$$

$$\pi = \tfrac{1}{4}(\Sigma y_{1j1} - \Sigma y_{1j2} + \Sigma y_{2j1} - \Sigma y_{2j2}) \sim N(\pi, m\sigma_\varepsilon^2/8)$$

$$\tau = \tfrac{1}{4}(\Sigma y_{1j1} - \Sigma y_{1j2} - \Sigma y_{2j1} + \Sigma y_{2j2}) \sim N(\tau - \tfrac{1}{2}\alpha, m\sigma_\varepsilon^2/8)$$

$$\alpha = \tfrac{1}{2}(\Sigma y_{1j1} + \Sigma y_{1j2} - \Sigma y_{2j1} - \Sigma y_{2j2}) \sim N(\alpha, m\sigma_\varepsilon^2/2)$$

plus quadratic forms SSP $\sim \sigma_\varepsilon^2\chi^2_{N-2}$

and SSE $\sim \sigma_\varepsilon^2\chi^2_{N-2}$ ...9.2.2

Prior distributions are postulated for μ, π, τ, α, σ_ε^2, and σ_β^2. As is usual when there is no strong a priori information about the parameters, uninformative priors are used except for the carry-over effect α, which is assumed to be normally distributed with mean 0 and variance σ_α^2. The joint prior density is thus:

$$p(\mu, \pi, \tau, \alpha, \sigma_\varepsilon^2, \sigma_\beta^2) \propto \sigma_\varepsilon^{-2}\sigma_\beta^{-2}\exp\{-\alpha/(2\sigma_\alpha^2)\} \quad ..9.2.3$$

By giving σ_α^2 different values, a range of prior beliefs about the likelihood of the presence of carry-over may be obtained. $\sigma_\alpha^2 = 0$ corresponds to a prior belief that carry-over is not possible, while $\sigma_\alpha^2 = \infty$ corresponds to an uninformative prior about the size of the carry-over effect.

Multiplying the prior distribution and the likelihood gives the joint posterior distribution. The parameters of most interest are τ and α, and the joint posterior distribution of these could be found if the other parameters could be integrated out of the joint posterior. Unfortunately, it is only possible to

integrate out μ and π analytically, with the other parameters having to be removed by the use of numerical integration. Thus, exact general results cannot be obtained for the joint distribution of τ and α, and each dataset would have to be submitted to numerical integration separately, using a package such as the Bayes 4 program, developed at the University of Nottingham. However, Grieve shows that approximations to the posterior marginal distributions of τ and α, and the conditional posterior distribution of τ given α are possible to obtain. These are given as:

$$p(\alpha \mid \mathbf{y}) = t(\hat{\alpha},\ m(SSP)/[2(N-2)],\ N-2)$$
$$p(\tau \mid \alpha,\ \mathbf{y}) = t(\hat{\tau} + \tfrac{1}{2}\alpha,\ m(SSE)/[8(N-2)],\ N-2)$$
$$p(\tau \mid \mathbf{y}) = t(\hat{\tau} + \tfrac{1}{2}\hat{\alpha},\ mf/(8h),\ f) \qquad \ldots 9.2.4$$

where $t(\theta, \phi, v)$ denotes a shifted and scaled t distribution with v degrees of freedom, location parameter θ, and scale parameter $\phi^{1/2}$; and

$f = 4 + \{(SSE + SSP)^2(N-6)\}/\{SSE^2 + SSP^2\}$

$h = (f-2)(SSE + SSP)/(N-4)$.

9.3 The Two-period Design with Baselines

In chapter 4 it was shown that the addition of baseline measurements to the two-period cross-over gave a more efficient estimator of the carry-over effect. However, from the point of view of a Bayesian analysis,

the addition of baseline observations makes the situation more complex, since the likelihood of a set of three observations on a single subject must now be considered as having a tri-variate normal distribution. Writing $\mathbf{y}$ for the vector of three observations on a single subject, i.e. $\mathbf{y}' = (y_{ij0}, y_{ij1}, y_{ij2})$, the tri-variate normal distribution has p.d.f.

$$(2\pi)^{-1/2}|\Sigma|^{-1/2}\exp\{-\tfrac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\} \qquad \ldots 9.3.1$$

where $\boldsymbol{\mu}$ is the vector of means given by the linear model, and Σ is the variance-covariance matrix of the three observations, i.e. $\Sigma = \sigma_T^2 \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$

where $\sigma_T^2 = \sigma_B^2 + \sigma_W^2$ and $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_W^2)$

Writing y_0, y_1, and y_2 for y_{ij0}, y_{ij1}, and y_{ij2}, and μ_0, μ_1, and μ_2 for the expected values of these observations given by the linear model, the exponent of the tri-variate normal is:

$$-1/[2\sigma_T^2(1-\rho)(1+2\rho)]\{(1+\rho)[(y_0-\mu_0)^2 + (y_1-\mu_1)^2 + (y_2-\mu_2)^2] - 2\rho[(y_0-\mu_0)(y_1-\mu_1) + (y_1-\mu_1)(y_2-\mu_2) + (y_2-\mu_2)(y_1-\mu_1)]\} \qquad \ldots 9.3.2$$

Given observations on three periods (baseline and two treatment periods), two contrasts between the periods

can be defined. For convenience, the usual contrasts for linear and quadratic trend are used, so that orthogonal effects π_L, and π_Q are defined, with $\pi_L = \pi_0 - \pi_2$, and $\pi_Q = \pi_0 - 2\pi_1 + \pi_2$. Effects $\gamma = \gamma_1 = -\gamma_2$, $\tau = \tau_A = -\tau_B$ and $\alpha = \alpha_A = -\alpha_B$ relating to sequence, treatment and carry-over are also defined, together with the estimators:

$$\hat{\mu} = 1/6n\{\Sigma y_{1j0} + \Sigma y_{1j1} + \Sigma y_{1j2} + \Sigma y_{2j0} + \Sigma y_{2j1} + \Sigma y_{2j2}\}$$

$$\hat{\gamma} = 1/2n\{\Sigma y_{1j0} - \Sigma y_{2j0}\}$$

$$\hat{\pi}_L = 1/2n\{\Sigma y_{1j0} - \Sigma y_{1j2} + \Sigma y_{2j0} - \Sigma y_{2j2}\}$$

$$\hat{\pi}_Q = 1/2n\{\Sigma y_{1j0} - 2\Sigma y_{1j1} + \Sigma y_{1j2} + \Sigma y_{2j0} - 2\Sigma y_{2j1} + \Sigma y_{2j2}\}$$

$$\hat{\tau} = 1/2n\{-\Sigma y_{1j0} + \Sigma y_{1j1} + \Sigma y_{2j0} - \Sigma y_{2j1}\}$$

$$\hat{\alpha} = 1/2n\{-2\Sigma y_{1j0} + \Sigma y_{1j1} + \Sigma y_{1j2} + 2\Sigma y_{2j0} - \Sigma y_{2j1} - \Sigma y_{2j2}\}$$

It can be shown that the likelihood partitions into components depending only on $\hat{\mu}$, $\hat{\pi}_L$, $\hat{\pi}_Q$ and $\hat{\gamma}$, $\hat{\tau}$, $\hat{\alpha}$, with

with $\hat{\mu} \sim N(\mu, \sigma^2(1+2\rho)/6n)$,

$\hat{\pi}_L \sim N(\pi_L, \sigma^2(1-\rho)/n)$

and $\hat{\pi}_Q \sim N(\pi_Q, 3\sigma^2(1-\rho)/n)$

The term involving $\hat{\gamma}$, $\hat{\tau}$, and $\hat{\alpha}$ cannot be factored, and has the form of a tri-variate normal, with mean vector

$$\mu_\gamma = (\Sigma y_{1j0} - \Sigma y_{2j0})/N$$

$$\mu_\tau = (-\Sigma y_{1j0} + \Sigma y_{1j1} + \Sigma y_{2j0} - \Sigma y_{2j1})/N$$

$$\mu_\alpha = (-2\Sigma y_{1j0} + \Sigma y_{1j1} + \Sigma y_{1j2} + 2\Sigma y_{2j0} - \Sigma y_{2j1} - \Sigma y_{2j2})/N$$

and variance covariance matrix

$$\sigma^2/N \begin{pmatrix} 1 & -(1-\rho) & -2(1-\rho) \\ -(1-\rho) & 2(1-\rho) & 3(1-\rho) \\ -2(1-\rho) & 3(1-\rho) & 6(1-\rho) \end{pmatrix}$$

It is interesting that the mean for τ in this trivariate normal is the contrast between the baseline and first period observations, the presence of the carry-over effect preventing the use of the data from the second period.

Assuming vague priors for the parameters as before, the posterior will be in a similar form to the above likelihood, presenting great difficulty in obtaining posterior distributions for the parameters of interest. Integrating out the parameters of little interest is not possible analytically, and, although numerical integration techniques might be applied, this requires sophisticated computing. This is an example of the problems associated with Bayesian analysis, which can often lead to intractable joint posterior distributions. There seems to be little point in pursuing this method of approach, as its application to more complicated designs is likely to yield still more intractable joint posteriors. Fortunately, an alternative approach utilising Gibbs sampling has now been suggested, which is much simpler to apply. The application of this method will now be considered.

9.4 Gibbs Sampling

It is evident that a straightforward Bayesian analysis often gives a joint posterior distribution from which the marginal distributions of interest cannot be derived analytically. In some cases approximate distributions can be obtained as above(Grieve[1989]), while in others numerical integration techniques would be needed (Naylor & Smith [1982,1988]). A simpler alternative to the use of numerical integration techniques is to employ a "sampling based approach". Gelfand & Smith[1990] and Gelfand et al[1990] have investigated the use of the Gibbs sampler, described by Geman & Geman[1984], for estimating the conditional posterior distributions in a Bayesian analysis, and are enthusiastic about its efficacy. The technique is, indeed, remarkably simple in essence. The conditional distribution of a parameter is calculated, assuming that all other unknown parameters are in fact known or fixed in value. A value of the one parameter which is being regarded as unknown at this stage is then simulated, taking a previously simulated value of the other parameters as their 'true' value. The simulated value of the parameter is then used in further simulations of the other parameters. Thus the technique cycles round the set of parameters simulating each one in turn. This cyclic process is repeated for many simulation sets, and Geman & Geman showed that under mild conditions the process would

converge, allowing the required marginal distribution to be estimated from the conditional distributions on the ultimate cycle.

Because of the notational difficulties, it is simpler at this stage to think in terms of a particular number of parameters θ, although it is not necessary to specify the nature of the parameters. We will consider a set of 3 parameters θ_1, θ_2 and θ_3. The conditional posterior for θ_1, given the observations $\mathbf{x}$, and considering θ_2 and θ_3 to be known, may be written $h(\theta_1 \mid \theta_2, \theta_3, \mathbf{x})$. Similar expressions for the conditional posteriors of θ_2 and θ_3 will also be used.

The Gibbs sampling procedure will start with initial values of θ_1, θ_2 and θ_3, denoted $\theta_i^{(0)}$. These initial values will be changed by simulating a new value at each cycle of the iterative process. The value for θ_i simulated on cycle k will thus be denoted $\theta_i^{(k)}$. The process of Gibbs sampling then proceeds as in fig 9 below.

Gelfand et al[1990] suggest that convergence of the process can be judged by plotting the ordered parameter values from successive cycles, which will lie approximately on a line of slope 1 when the process has converged. Assuming that the process has converged after k cycles, the marginal posterior distribution can be estimated using the $\theta^{(k)}$ values from the simulation sets. Let the values for the ith simulation set be denoted θ_{1i}, θ_{2i} and θ_{3i}. In order to estimate the marginal posterior

for θ_1, we will use the fact that the form of the conditional posterior of θ_1, given a value for θ_2 and θ_3, is known, ($h(\theta_1 \mid \theta_2, \theta_3, \mathbf{x})$). Using θ_{21}, θ_{31} as the "known" values of θ_2 and θ_3, we can calculate ordinates of the posterior for θ_1. Because the values of θ_{21} and θ_{31} will be slightly different for each simulation set, the calculated ordinate for a particular value of θ_1 will vary. The overall estimate of the ordinate of the marginal posterior of θ_1 is taken as the average of the estimates from the individual simulation sets.

Fig 9 : Sequence of Operations in Gibbs Sampling

Start cycle k with values $\theta_1^{(k-1)}$, $\theta_2^{(k-1)}$, $\theta_3^{(k-1)}$

Simulate a value from $h(\theta_1 \mid \theta_2^{(k-1)}, \theta_3^{(k-1)}, \mathbf{x})$

- simulated value is $\theta_1^{(k)}$

Simulate a value from $h(\theta_2 \mid \theta_1^{(k)}, \theta_3^{(k-1)}, \mathbf{x})$

- simulated value is $\theta_2^{(k)}$

Simulate a value from $h(\theta_3 \mid \theta_1^{(k)}, \theta_2^{(k)}, \mathbf{x})$

- simulated value is $\theta_3^{(k)}$

In order to illustrate the technique, its application to a very simple situation, for which the marginal posterior distributions can be derived

analytically, will be considered, before the technique is applied to the more complex situations which arise in the analysis of cross-over designs.

9.5 A Simple Example

Consider a situation in which a random sample of n observation is drawn from a normal population whose mean and variance are unknown. The two unknown parameters, the population mean, and the population variance will be denoted θ_1 and θ_2 respectively, while the sample observations will be denoted $x_1, \ldots, x_n$. The likelihood is then given by

$$L = (2\Pi\theta_2)^{-\frac{1}{2}n} \exp\{ -\Sigma(x_i - \theta_1)^2/2\theta_2\}$$

If independent vague priors are specified for the two parameters, with a uniform prior for θ_1 and a prior for θ_2 proportional to $1/\theta_2$, the joint posterior is proportional to

$$(\theta_2)^{-\frac{1}{2}(n+2)} \exp\{ -\Sigma(x_i - \theta_1)^2/2\theta_2\}$$

Writing $\Sigma(x_i - \theta_1)^2$ as $(n-1)s^2 + n(\bar{x} - \theta_1)^2$, where $\bar{x}$ is the mean of the sample and s^2 the unbiased estimator of the population variance from the sample, the joint posterior can be re-written

$$(\theta_2)^{-\frac{1}{2}(n+2)} \exp\{ -[(n-1)s^2 + n(\bar{x} - \theta_1)^2]/2\theta_2\}$$

Treating θ_2 as a known constant, the conditional posterior for θ_1 is

$$p(\theta_1|\theta_2, \mathbf{x}) \alpha \exp\{ -n(\bar{x} - \theta_1)^2/2\theta_2\}$$

multiplying this by the constant $(\theta_2)^{-n}$, normalises the expression for this conditional distribution, showing it to be the normal distribution with mean $\bar{x}$, and variance θ_2/n. If θ_1 is treated as a constant, the conditional posterior for θ_2 can be obtained

$$p(\theta_2|\theta_1, \mathbf{x}) \propto (\theta_2)^{-n(n+2)} \exp\{ -[(n-1)s^2 + n(\bar{x} - \theta_1)^2]/2\theta_2\}$$

This indicates that $[(n-1)s^2 + n(\bar{x} - \theta_1)^2]/\theta_2$ has a chi-squared distribution with n degrees of freedom.

Having derived the conditional posterior distributions, the Gibbs sampling can be instituted. Adapting the general method given in fig 9 to this specific example, gives the sequence of steps illustrated in fig 10

Fig 10 : Gibbs Sampler for the Simple Example

Start cycle k with values $\theta_1^{(k-1)}$, $\theta_2^{(k-1)}$

Simulate a value from $N(\bar{x}, \theta_2^{(k-1)}/n)$

- simulated value is $\theta_1^{(k)}$

Simulate a value from the chi-squared distribution with n degrees of freedom = C_n

- $[(n-1)s^2 + n(\bar{x} - \theta_1^{(k)})^2]/C_n$ is $\theta_2^{(k)}$

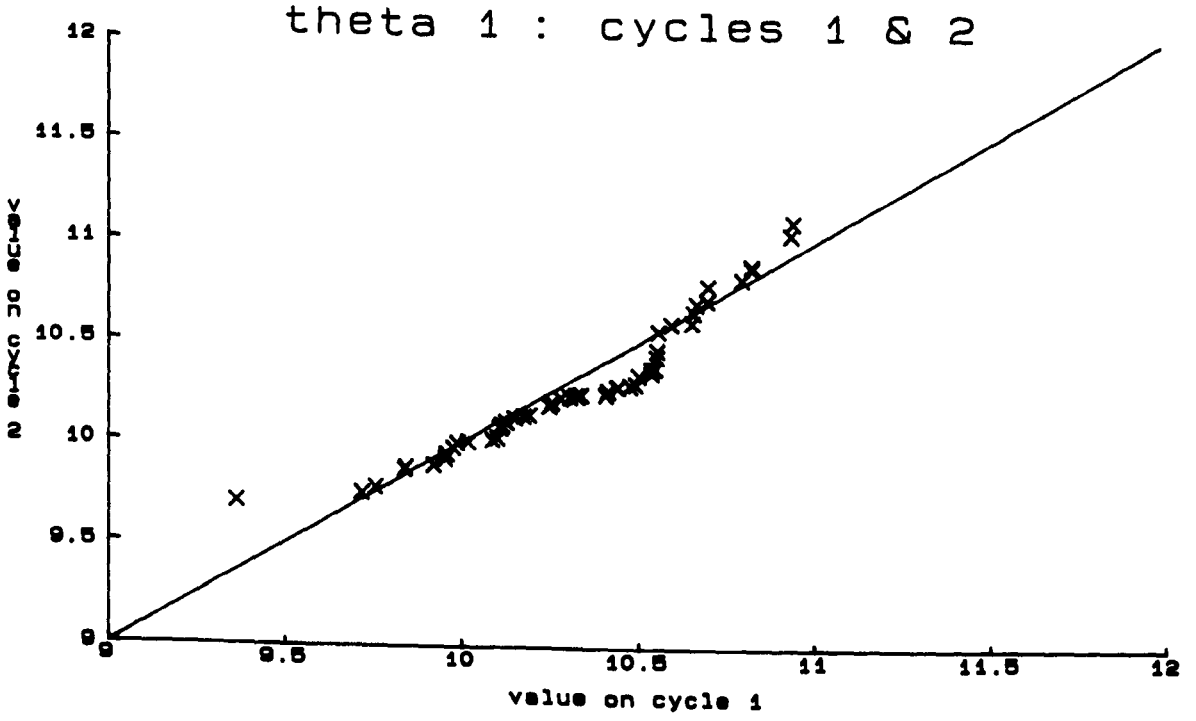
Assuming that the process has converged after k cycles, the $\theta^{(k)}$ values for the ith simulation set will

be denoted θ_{1i} and θ_{2i}. The conditional posterior distribution for θ_1 given θ_{2i} is a normal distribution with mean $\bar{x}$, and variance θ_{2i}/n. Let the ordinate of this distribution be denoted $h(\theta_1|\theta_{2i}, \mathbf{x}) \sim N(\theta_1, \theta_{2i}/n)$. Such an ordinate can be found for each simulation set, and the average of the ordinates given by all the simulation sets is used as an estimate of the corresponding ordinate of the posterior distribution of θ_1,
ie $f(\hat{\theta}_1|\mathbf{x}) = \Sigma h(\theta_1|\theta_{2i}, \mathbf{x})/N$, where N is the number of simulation sets. This allows us to plot the shape of the marginal posterior distribution, and calculate approximate "confidence" intervals for θ_1.

For a simple situation such as this, convergence is achieved quite quickly, with about 25 cycles proving sufficient. The ordered values of θ_1 and θ_2, from which convergence can be judged, are plotted for cycles 1 & 2, 12 & 13 and 24 & 25 in figs. 11, 12 & 13

In this simple case, the form of the marginal posteriors is easily obtained analytically, (see eg Lindley[1965]), the marginal posterior for θ_1 being such that $n^{\frac{1}{2}}(\theta_1 - \bar{x})/s$ has a t-distribution with n-1 degrees of freedom, with the marginal posterior for θ_2 being such that $(n-1)s^2/\theta_2$ has a chi-squared distribution with n-1 degrees of freedom. A program was written to perform the Gibbs sampling proceedure for this simple example, using NAG routines to simulate values from the required normal and chi-squared distributions. The estimated ordinates of

Fig 11 : Convergence of the Gibbs Sampler for the Simple Example : Cycles 1 & 2



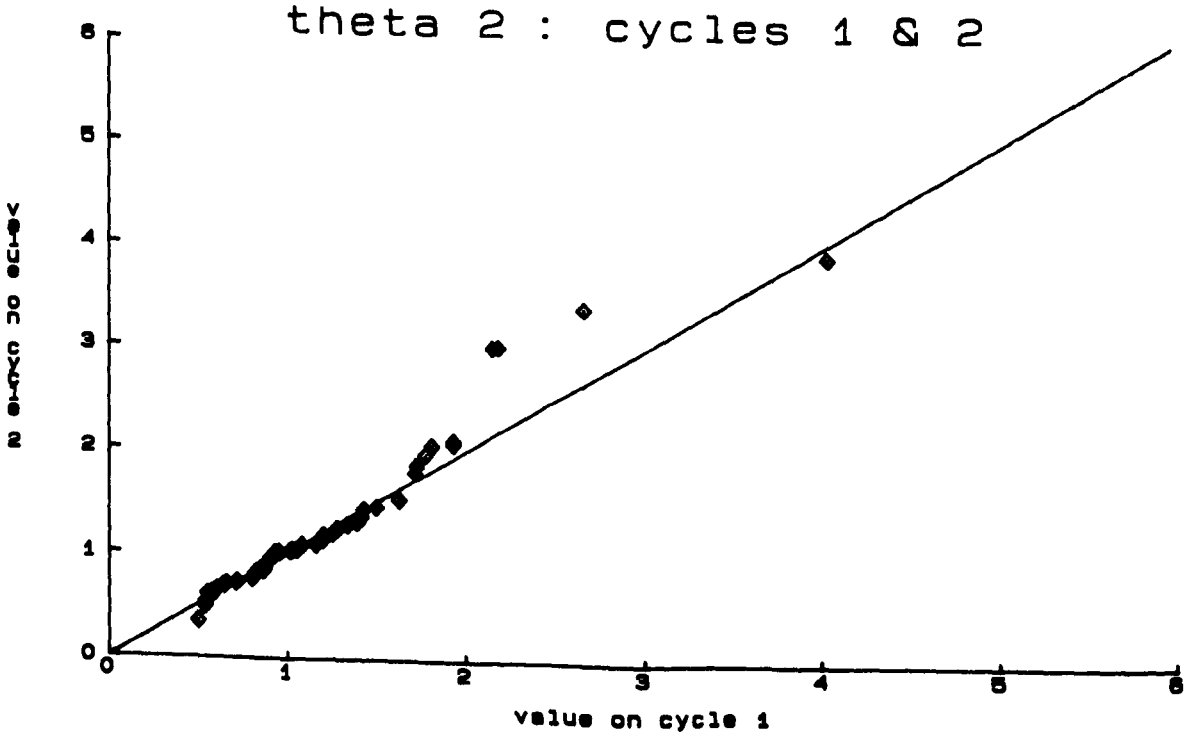
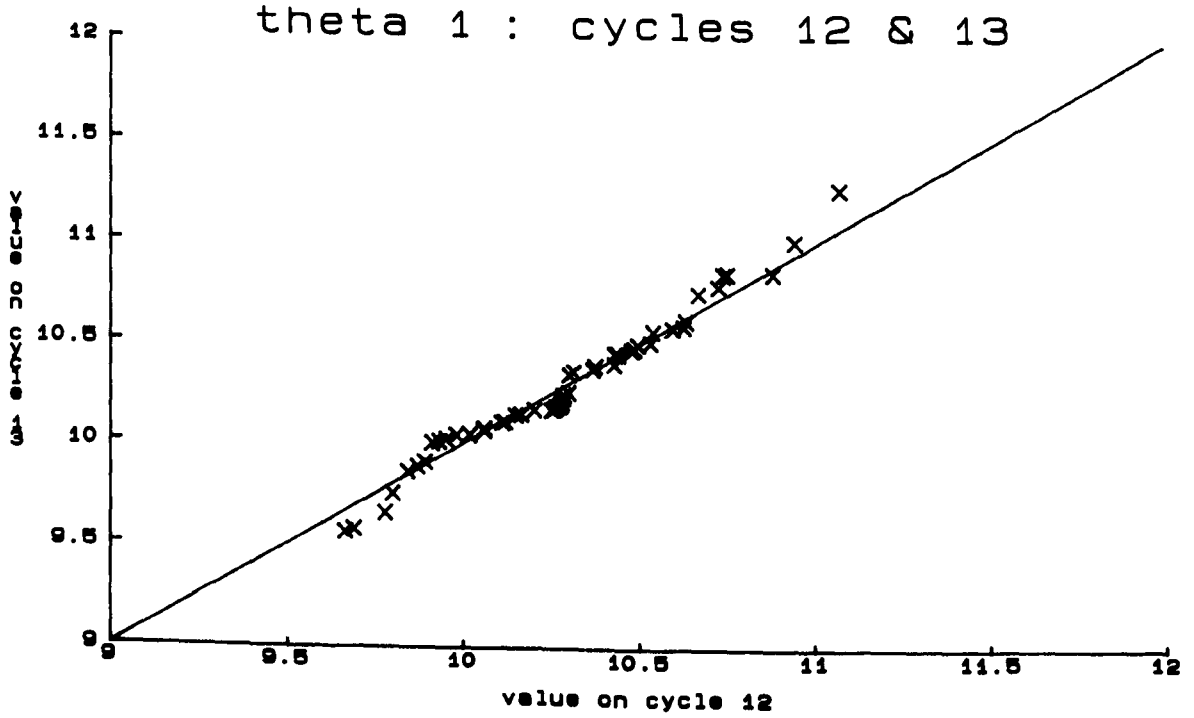


Fig 12 : Convergence of the Gibbs Sampler for the Simple Example : Cycles 12 & 13



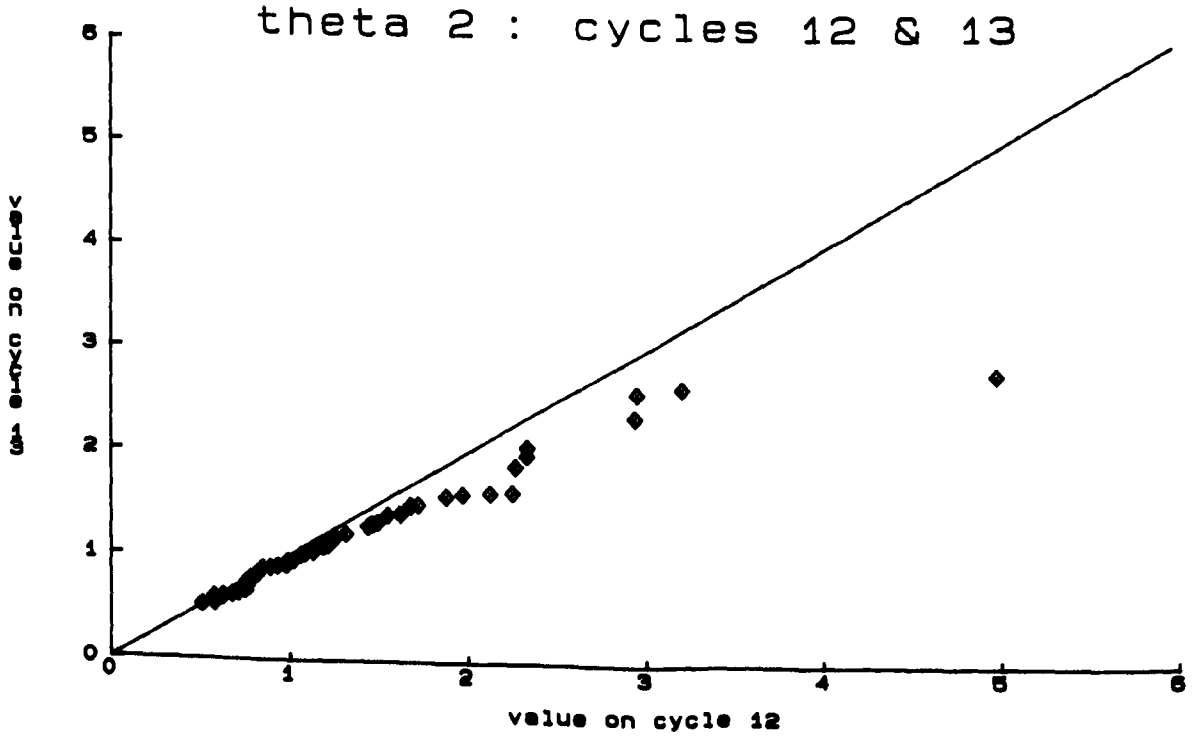
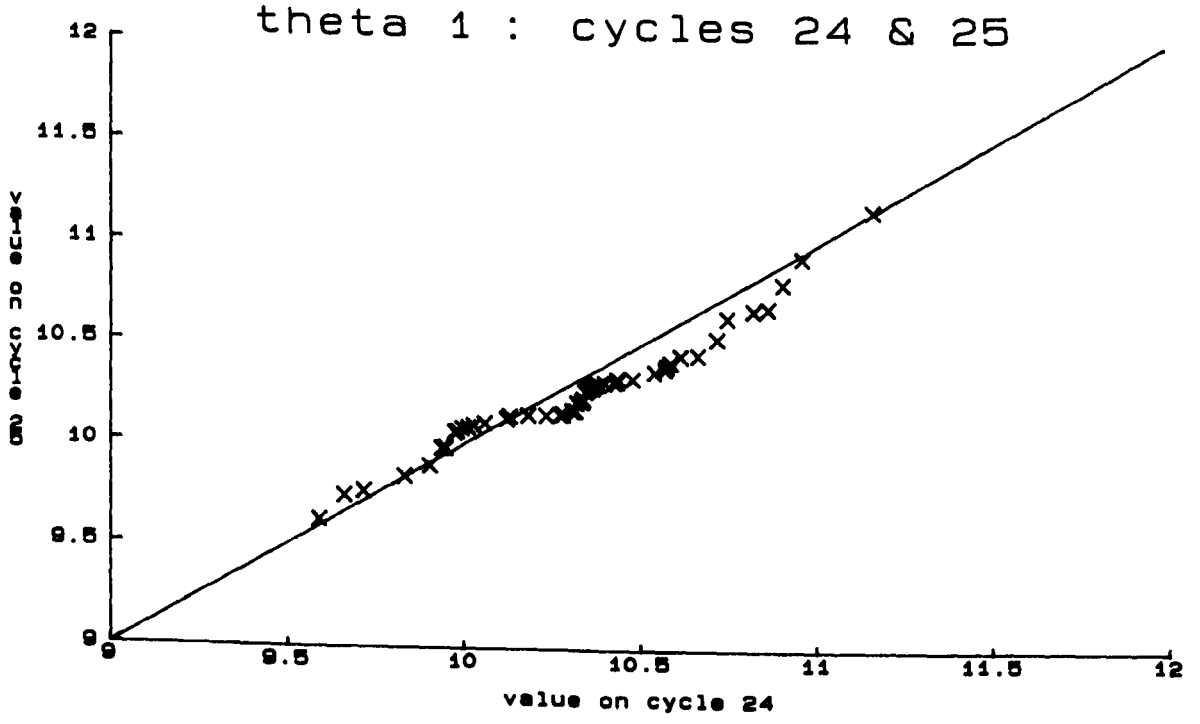


Fig 13 : Convergence of the Gibbs Sampler for the Simple Example : Cycles 24 & 25



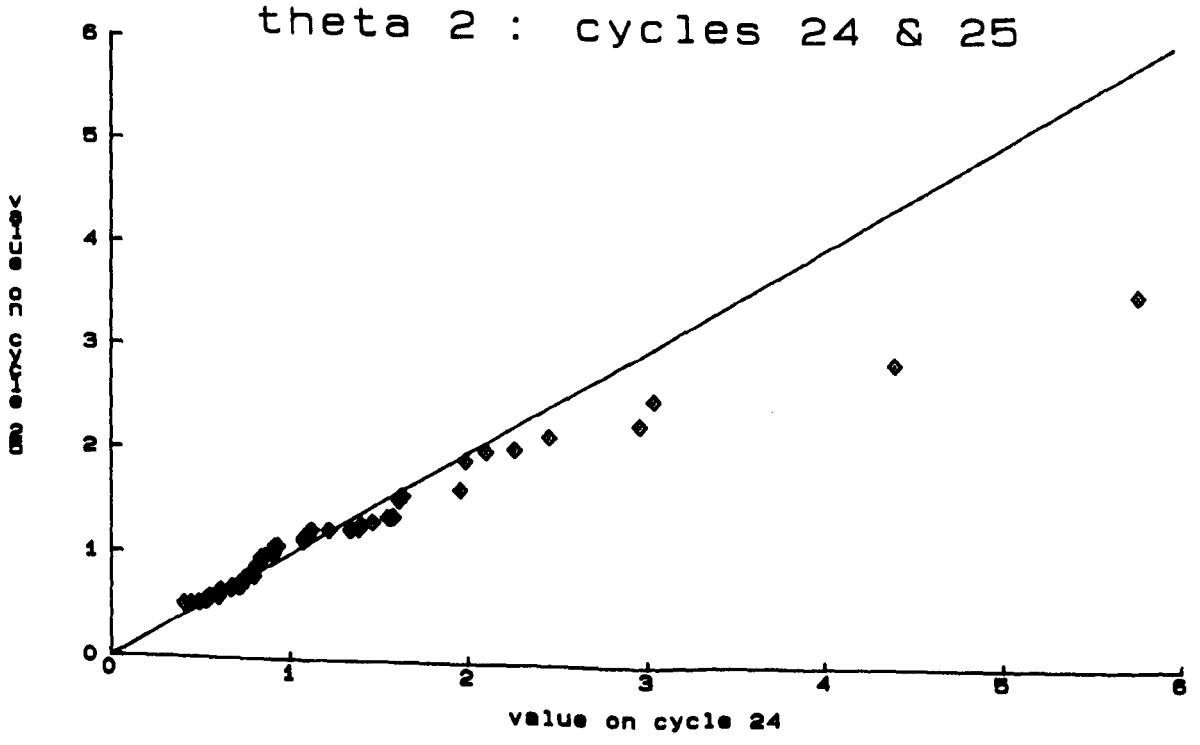
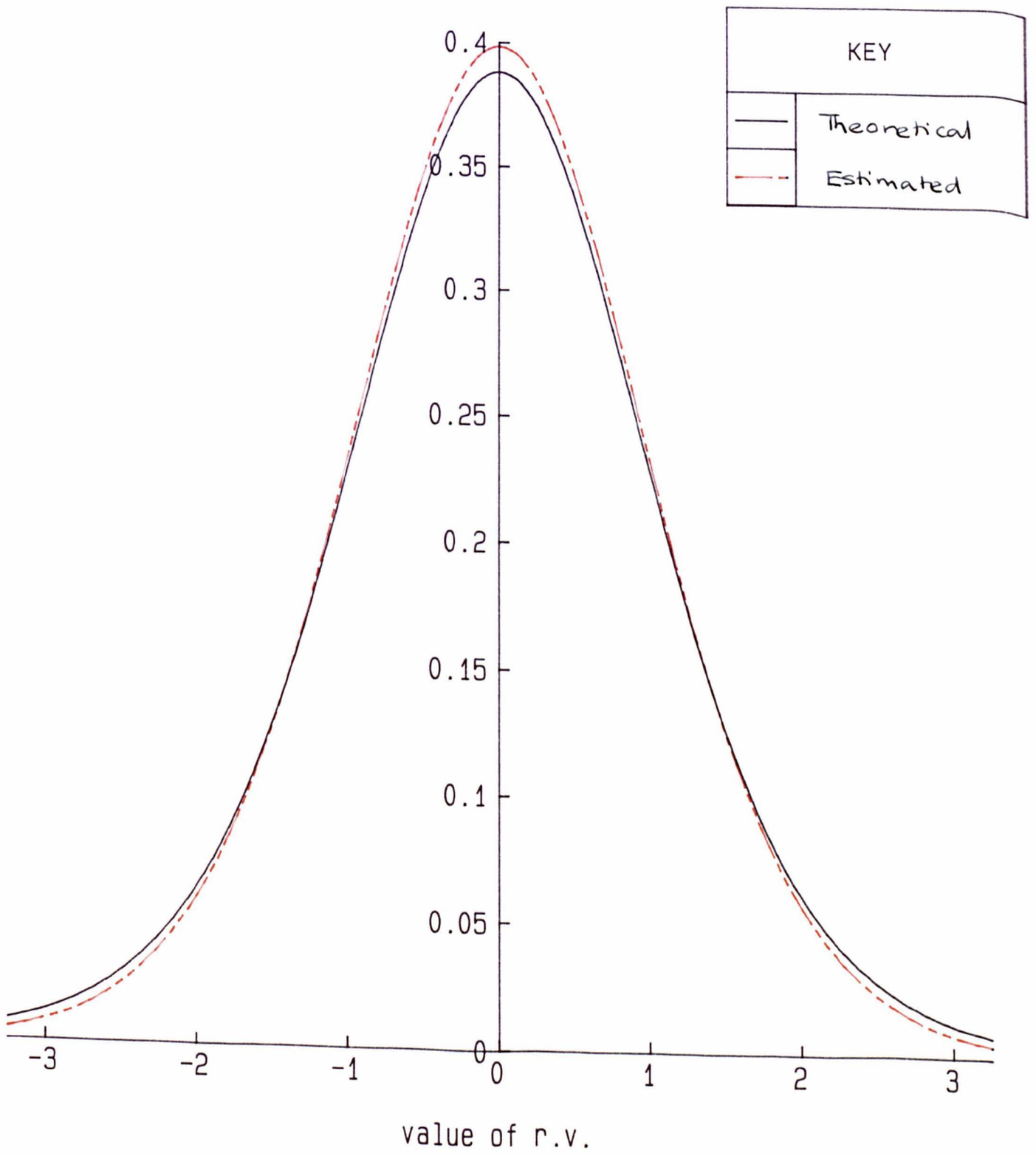


Fig 14 : Estimated and Theoretical Posteriors for θ_1



the marginal posterior distribution of θ_1, together with the t-distribution which is the theoretical form of the posterior are plotted in fig 14, showing good agreement between the Gibbs sampling and theoretical results.

9.6 Gibbs Sampling for the Two-period Cross-over

The analysis of the two-period cross-over using Gibbs sampling described here follows the method given by Gelfand et al[1990], who give several examples of the application of Gibbs sampling, including the analysis of a two-period cross-over with missing values, but without carry-over. Initially, details of the analysis with no carry-over in the model will be given, and this will then be generalised to include carry-over. The methods are then used to analyse the enuresis data given by Hills & Armitage[1979], which has been extensively investigated using classical methods (Poloniecki & Daniel[1981]), and was also used by Grieve[1985] as an example in his Bayesian analysis. A listing of the data is given in appendix F.

The linear model without carry-over will be as given in chapter 2, except that the constraints on the parameters have been incorporated so that, for the period effects, $\pi = \pi_1 = -\pi_2$, and for the treatment effects, $\tau = \tau_A = -\tau_B$. As before, the parameters μ and ω represent the overall mean and the subject effect, respectively, with ϵ representing the error term. The model for the response

for the jth subject in the ith sequence for the kth period is thus:

$$y_{ijk} = \mu + (-1)^{k-1}\pi + (-1)^{i+k}\tau + \omega_{ij} + \epsilon_{ijk}$$

ω and ϵ are random effects with variances σ_ω^2 and σ_ϵ^2, respectively. The two observations on a single subject therefore have a bivariate normal distribution with mean $\mathbf{X}_{ij}\boldsymbol{\theta}$, and variance-covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\theta}' = [\mu, \pi, \tau]$, the vector of model parameters, and $\mathbf{X}_{ij}$ is the appropriate design matrix. i.e.

$$\mathbf{X}_{1j} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \qquad \mathbf{X}_{2j} = \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\omega^2+\sigma_\epsilon^2 & \sigma_\omega^2 \\ \sigma_\omega^2 & \sigma_\omega^2+\sigma_\epsilon^2 \end{pmatrix}$ being the same for all subjects

Two new variables, $y_{\bar{i}j}$ and $y_{\hat{i}j}$ are defined as follows:

$$y_{\bar{i}j} = \tfrac{1}{2}(y_{ij1} - y_{ij2})$$

$$y_{\hat{i}j} = \tfrac{1}{2}(y_{ij1} + y_{ij2})$$

These new variables have the advantage of being orthogonal, so that the bivariate normal distribution of $y_{\bar{i}j}$ and $y_{\hat{i}j}$ is relatively simple. Thus

$$\begin{pmatrix} y_{\bar{1}j} \\ y_{\hat{1}j} \end{pmatrix} \sim N \begin{pmatrix} \pi+\tau \\ \mu \end{pmatrix} \tfrac{1}{2}\begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} y_{\bar{2}j} \\ y_{\hat{2}j} \end{pmatrix} \sim N \begin{pmatrix} \pi-\tau \\ \mu \end{pmatrix} \tfrac{1}{2}\begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

where $\sigma_2^2 = 2\sigma_\omega^2 + \sigma_\epsilon^2$

Prior distributions are now specified for $\boldsymbol{\theta}$, σ_ϵ^2 and σ_2^2. $\boldsymbol{\theta}$ is assumed to have a trivariate normal distribution, and the two variances are assumed to have inverse gamma distributions. The distributions are

assumed to be independent, so that the prior distribution is given by:

$$f(\theta, \sigma_1^2, \sigma_2^2) = N(M, C)\, IG(\tfrac{1}{2}d_e, \tfrac{1}{2}d_e v_e)\, IG(\tfrac{1}{2}d_2, \tfrac{1}{2}d_2 v_2)$$

This distribution is subject to the restriction $\sigma_1^2 \leq \sigma_2^2$, which is a consequence of the definition of σ_2^2.

Defining $SS_1 = 2\Sigma\{y_{1j} - (\pi+\tau)/2)\}^2 + 2\Sigma\{y_{2j} - (\pi-\tau)/2\}^2$

and $SS_2 = 2\Sigma\Sigma\{y_{ij} - \mu\}^2$

with $X'S^{-1}Y = \Sigma\Sigma X_{ij}'\Sigma^{-1}Y_{ij}$, $X'S^{-1}X = \Sigma\Sigma X_{ij}'\Sigma^{-1}X_{ij}$

and $D^{-1} = X'S^{-1}X + C^{-1}$, it can be shown that the Gibbs sampler for σ_1^2, σ_2^2, and θ is specified by:

$$\sigma_1^2 | Y, \theta, \sigma_2^2 = IG(\tfrac{1}{2}(n+d_e),\ \tfrac{1}{2}(SS_1 + d_e v_e)) \quad (\sigma_1^2 \leq \sigma_2^2)$$

$$\sigma_2^2 | Y, \theta, \sigma_1^2 = IG(\tfrac{1}{2}(n+d_2),\ \tfrac{1}{2}(SS_2 + d_2 v_2)) \quad (\sigma_1^2 \leq \sigma_2^2)$$

$$\theta | Y, \sigma_1^2, \sigma_2^2 = N\{D(X'S^{-1}Y + C^{-1}M), D\}$$

Details of this derivation are given in appendix D.

The Gibbs sampling process for this example is illustrated below in fig 15. With no carry-over in the model, the process converged quickly, and was relatively insensitive to different initial estimates of the parameters. The obvious initial estimates are those derived from the usual analysis given by Hills and Armitage[1979], and quoted in chapter 2 above. The estimate of the posterior distribution of the treatment effect obtained agrees with that given by Grieve when carry-over is assumed to be zero. A plot of the estimated posterior, is given in fig 16 together with the estimated posteriors when there is carry-over in the model.

Fig 15 : Gibbs Sampler for the Two-period Cross-over

Calculate SS_1 & SS_2 using current value of θ

Simulate a value from the gamma distibution with parameters $n+d_e$ and $SS_1+d_ev_e$. The new value of σ_1^2 is the reciprocal of this simulated value.

Simulate a value from the gamma distribution with parameters $n+d_2$ and $SS_2+d_2v_2$. The new value of σ_2^2 is the reciprocal of this simulated value.

Calculate D using the new values of σ_1^2 and σ_2^2

Simulate a value from the tri-variate normal distribution with mean $D(X'S^{-1}Y+C^{-1}M)$, and variance D. Simulated vector is new value of θ

9.7 Adding Carry-over to the Model

The addition of carry-over to the above model and the application of the Gibbs sampling approach is now considered. The overall form of the Gibbs sampler is unchanged, but θ is now a vector of four parameters, μ, π, τ and α, where α is the carry-over effect, with $\alpha_A = -\alpha_B = \alpha$. Details of the derivation of the conditional distributions are given in appendix D. Adding carry-over to the model increases the uncertainty about the size of

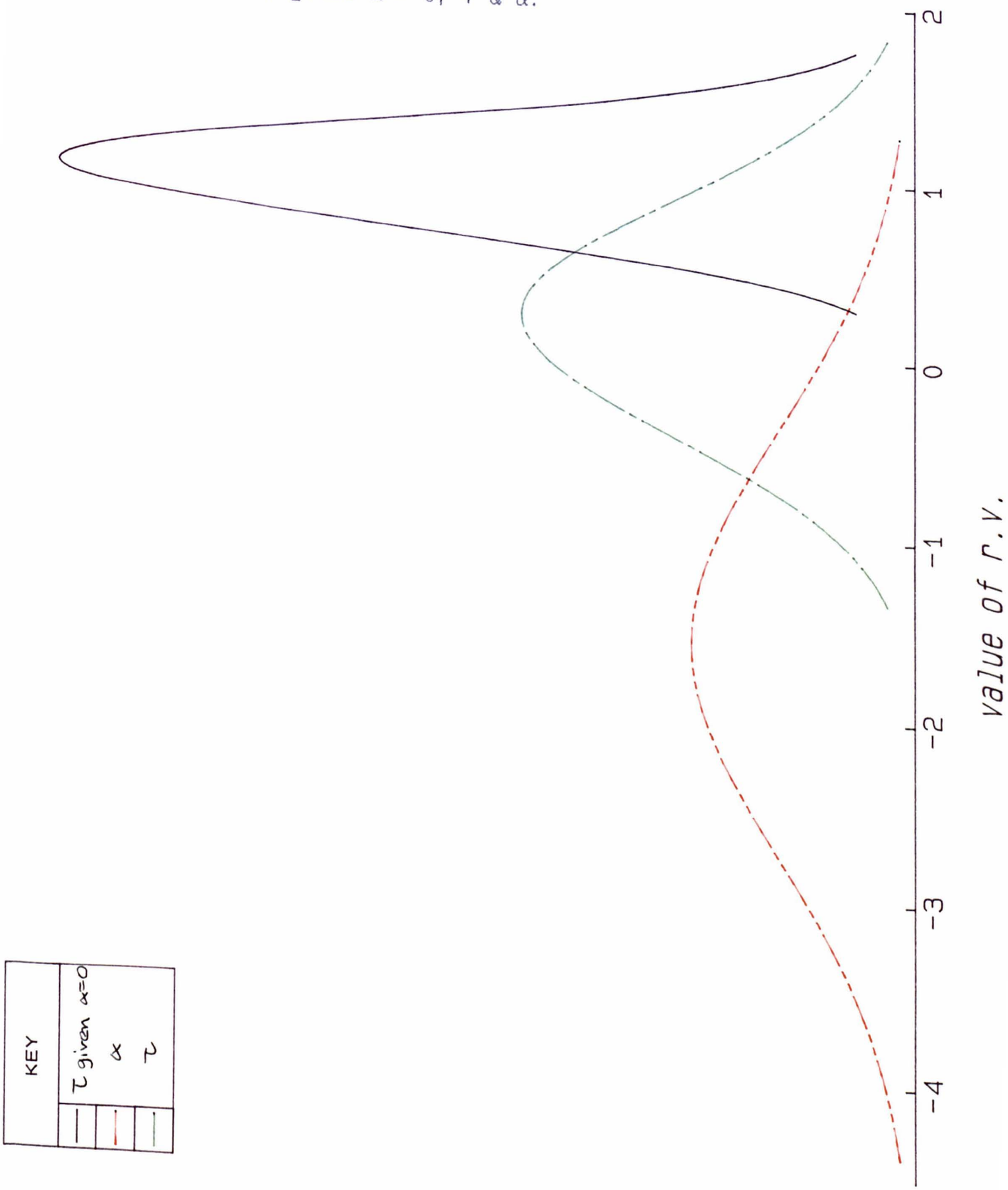
the effects, and the prior variances needed to be increased from the value used when there was no carry-over in order to allow the sampler to move to its optimum solution. The estimated posteriors for τ and α are plotted in fig 16, and agree very well with the results given in Grieve's paper, where an almost identical plot is given.

9.3 Three-period Designs

The sequence of operations in Gibbs sampling given in fig 9 could be applied to any experimental design, including any of the cross-over designs described previously. For each design it is necessary to derive the conditional distributions. Following the lead given in Gelfand et al[1990], this is simpler in terms of orthogonal functions of the observations on each subject, rather than the observations themselves. For the three-period design with sequences ABB and BAA, the orthogonal functions $2y_{ij1} - y_{ij2} - y_{ij3}$, which is used to obtain estimates of the treatment effect in the conventional analysis, and $y_{ij2} - y_{ij3}$, used to obtain estimates of the first-order carry-over effect, together with the total of the three observations on a subject $(y_{ij1} + y_{ij2} + y_{ij3})$ are those which seem most appropriate. These have been used to derive expressions for the conditional distributions, and details are given in appendix E. The vector of model parameters θ can be increased to include a group effect

Fig 16 : Estimated posterior for the enuresis data.

τ, given $\alpha = 0$; τ & α.



(γ), two period effects (π_1 & π_2), and, if required, first and second-order carry-over effect (α, β). The Gibbs sampler for σ_1^2, σ_3^2, and θ, where $\sigma_3^2 = 3\sigma_6^2 + \sigma_1^2$ that results is essentially the same in character as that described for the two-period design, with simulated values for σ_1^2 and σ_3^2 being obtained from inverse gamma distributions, and simulated values of θ being obtained from a multivariate normal distribution.

A FORTRAN program was written to apply the Gibbs sampler, both with and without second-order carry-over in the model. This was applied to the data for the ABB design, from Ebbutt[1984].

9.8.1 Results

In his paper on three-period cross-over designs, Ebbutt[1984] considers a set of data from a trial comparing two treatments for hypertension. The trial consisted of four sequences, ABB, BAA, ABA & BAB, but only the data from the ABB & BAA sequences have been used here. Systolic and diastolic blood pressure was recorded for each patient in each treatment period, and a baseline observation was also taken. Only the systolic blood pressures have been used here, and a listing of these is given in appendix F. Using the data from the three treatment periods, estimates of τ, α and β are 2.96, 1.336 and 8.116, repectively. The estimates of τ and α are obtained using the within-patient comparisons H_{ij} =

$2y_{ij1} - y_{ij2} - y_{ij3}$ and $K_{ij} = y_{ij2} - y_{ij3}$ respectively, and would be expected to be more accurate than the estimate for β which uses the between-patient comparison $2y_{ij1} + y_{ij2} + y_{ij3}$. However, the estimators of τ and α are only unbiased if $\beta = 0$. The variance of the estimator of β is large, and in a formal test, the hypothesis $\beta = 0$ would not be rejected. A better estimator of β can be obtained by using the baseline observations, and this gives $\beta = -2.86$. Corresponding estimates for τ and α are 2.246 and -0.094 respectively. The Gibbs sampler was run twice using each of these two sets of estimates, and the estimated posterior distributions are given in figs 17, 18, 19 & 20. Figs 17 and 19 give the estimated posteriors for τ and α when there is no second-order carry-over in the model. It can be seen that the different starting values make very little difference to these results. Figs 18 and 20 give the estimated posteriors for τ, α and β, and show a marked difference, both from the posteriors when there is no second-order carry-over, and from each other. Clearly, adding second-order carry-over to the model introduces a great deal more uncertainty, and this necessitates an increase in the prior variances. Although the program was run several times with increasing values of the prior variances until the solution appeared to stabilise, it may be that the optimum solution had not, in fact, been reached, and that the process had temporarily stabilised at a sub-optimum. Certainly, the

Fig 17 : Estimated posterior distribution for τ & α, given $\beta = 0$. Initial estimates $\tau = 2.96$, $\alpha = 1.336$

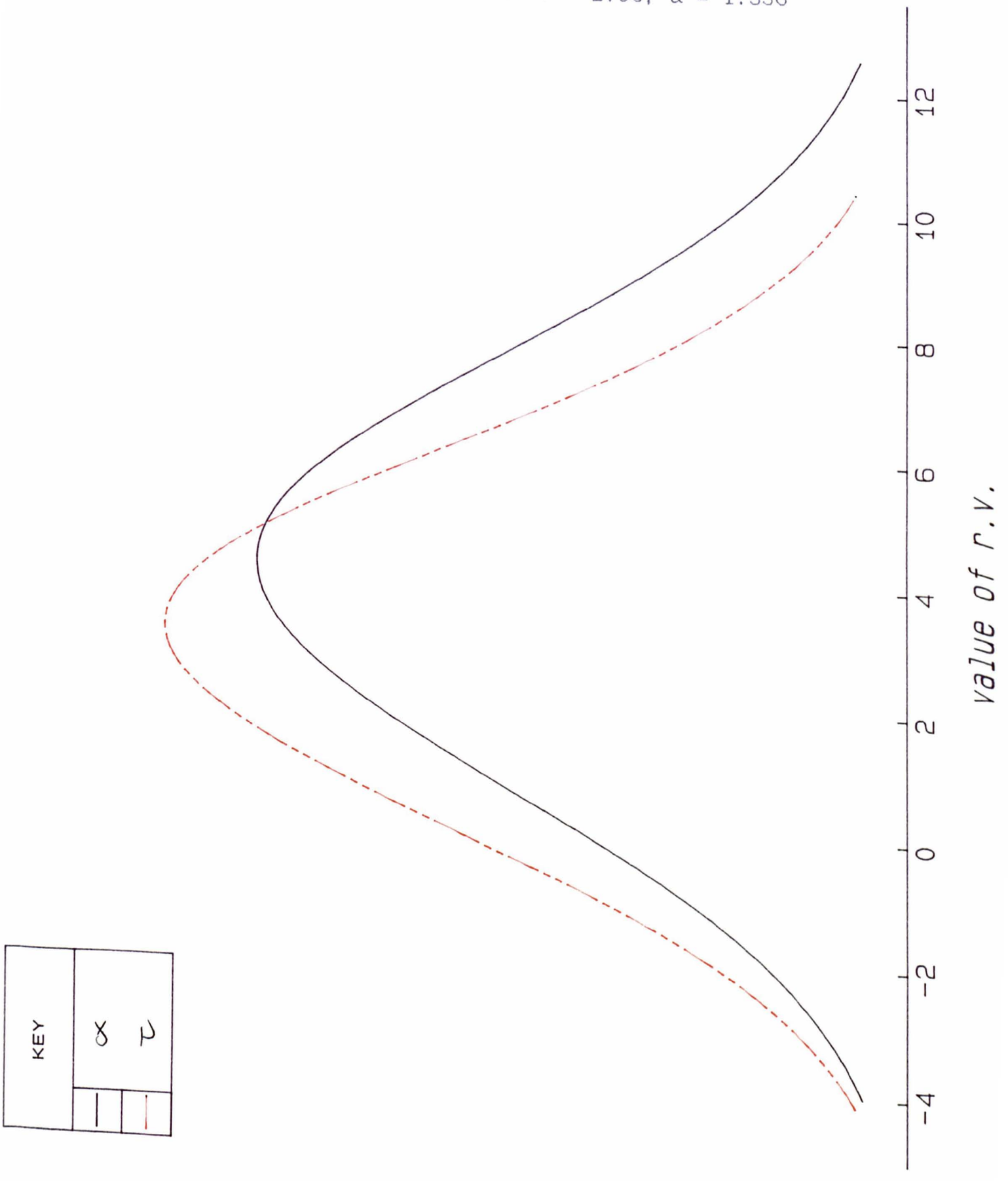


Fig 18 : Estimated posterior distribution for τ, α & β.

Initial estimates $\tau = 2.96$, $\alpha = 1.336$, $\beta = 8.116$

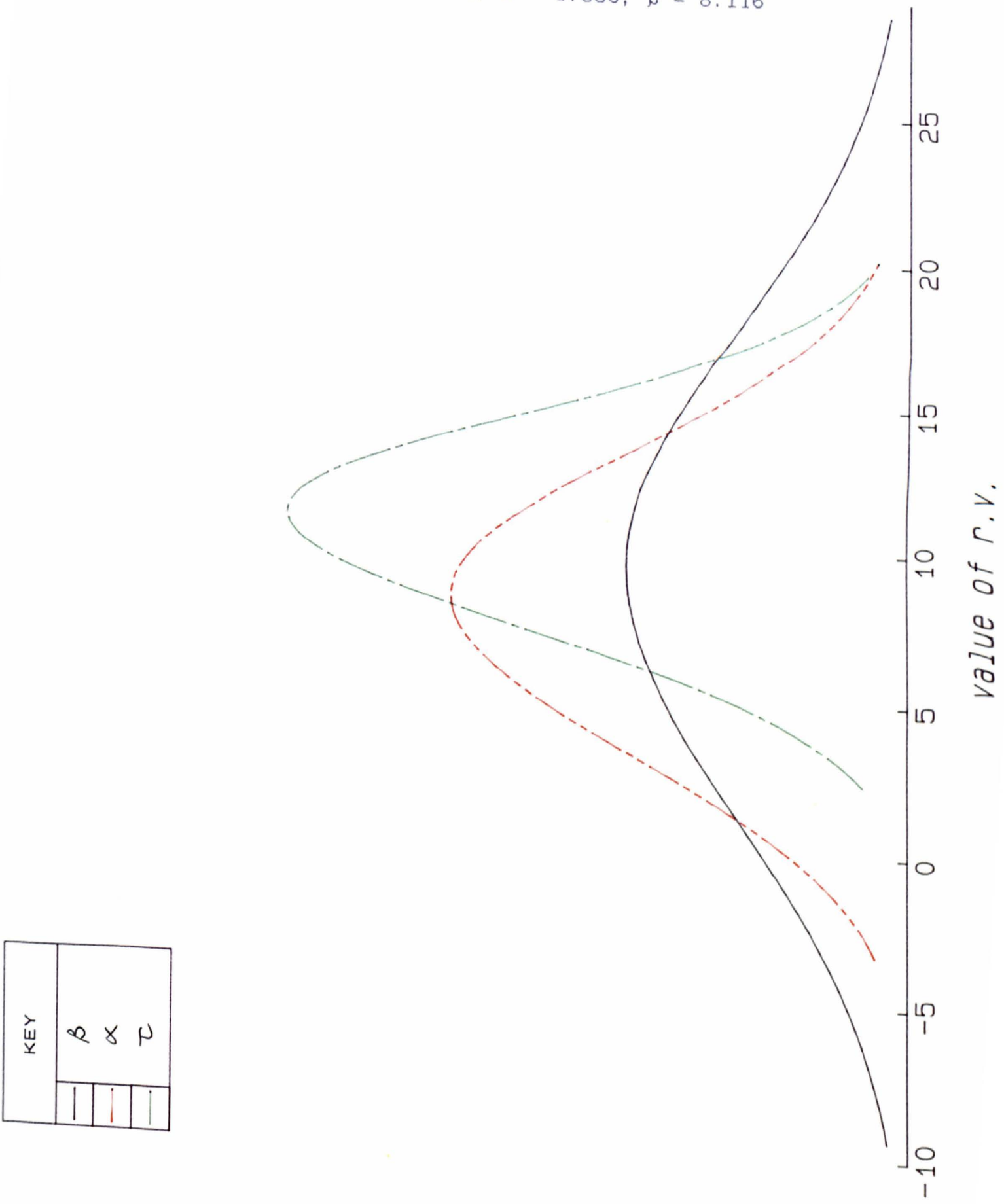


Fig 19 : Estimated posterior distribution for τ & α, given $\beta = 0$. Initial estimates $\tau = 2.246$, $\alpha = -0.094$

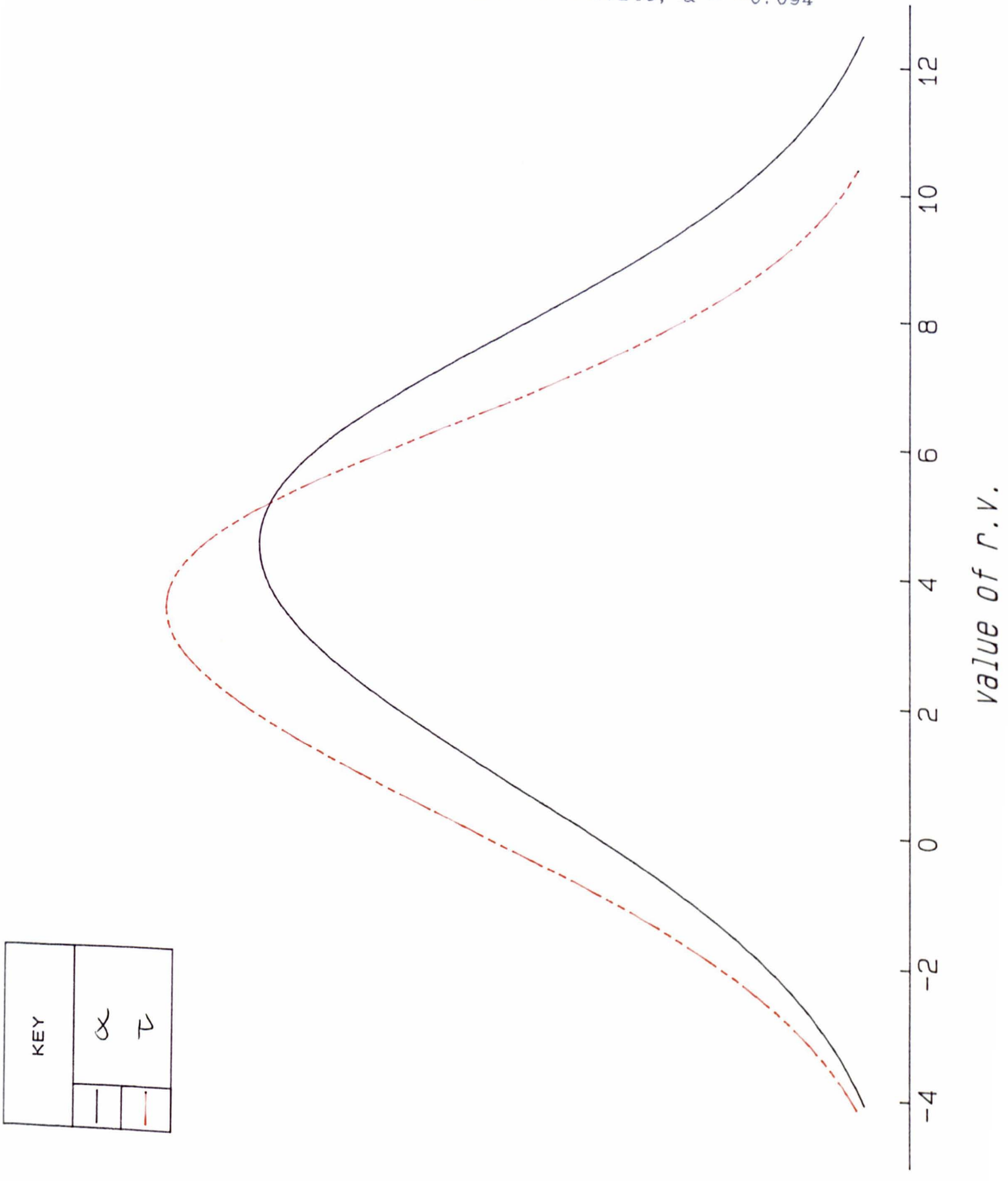
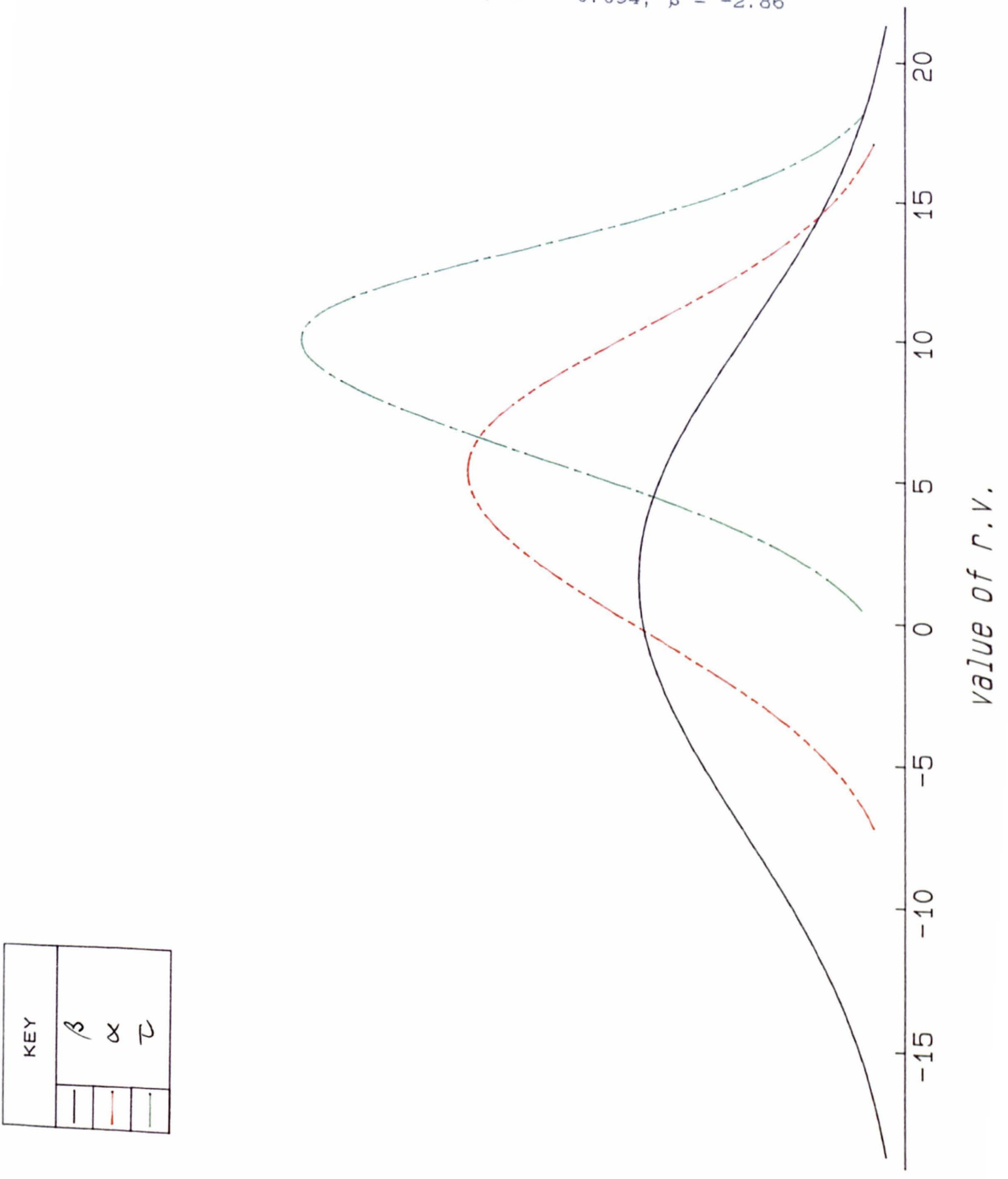


Fig 20 : Estimated posterior distribution for τ, α & β.

Initial estimates $\tau = 2.246$, $\alpha = -0.094$, $\beta = -2.86$



posteriors for β are centred at values relatively close to the initial estimate, indicating that the process has not moved far. On the other hand, the centres of the posteriors for τ and α are not particularly close to their starting values, and the estimates of τ are fairly similar. This may just reflect the fact that the data gives most information about τ and least about β, and suggests that the process had not properly converged.

9.9 Effectiveness of Gibbs Sampling

Gibbs sampling was found to work well for the two-period cross-over, and be relatively simple to apply. Using a model without carry-over, the computations were completed quickly, with rapid convergence, unaffected by the initial parameter estimates. Once carry-over was included in the model, the computations took much longer, and convergence could be adversely affected by inappropriate initial estimates of the parameters. However, it was found that use of the "obvious" initial estimates, from the Hills & Armitage estimators, generally avoided problems with convergence, and there is every reason to believe that the process will work well when such initial estimates are employed. Another critical factor in the success of the procedure was the choice of the variances of the prior distributions for the parameters. If these are too small, then the process is unable to move far from the initial estimates, while

if they are too large, the extra variability causes the condition that one of the simulated variances be larger than the other (a consequence of the fact that the second variance is the sum of the first and other variance terms) to be violated frequently by the simulated values. When this happens, another attempt at simulating suitable values must be made, and the larger the prior variances, the more likely it is that the simulated variances will violate this condition, necessitating more passes through the simulation loop to obtain acceptable values. This increases the computing time, which can quickly become unacceptably long. In some ways these two problems are related, because "good" initial estimates will be close to the optimal, and will not require the solution to move far. It is clearly good practice, however to gradually increase the prior variances and confirm that the solution remains stable.

It should not be thought that the Bayesian analysis overcomes all the difficulties of the classical analysis. Rather it is that, with a posterior distribution for the treatment effect, the uncertainty about its value is included, rather than quoting a, possibly misleading, point estimate. If some measure of the centre of the posterior distribution were to be quoted on its own, this would be as inadequate as the classical point estimate.

9.10 Discussion

Bayesian analysis has been suggested as a way out of the problems caused by carry-over, because it provides an overall picture, in the form of the posterior distributions, of the position regarding the estimation of the treatment and carry-over effects. It is clear, however, that the difficulties with integrating out unwanted parameters to obtain the required posteriors make the direct application of Bayesian methods very difficult. The method of Gibbs sampling overcomes these problems by obtaining an estimate of the posterior distributions indirectly, using a sampling approach. It is relatively easy to apply to complicated designs, and is not too expensive in computing time, given the powerful machines that are currently available. As with all simulation methods, however, there is a fine line between the method working well and providing a good estimate, and between it failing to converge, or performing badly. It appears that the choice of the initial estimates, and the prior variances are the important factors in this. In this investigation it has been found that the estimates from classical methods (e.g. Hills & Armitage for the two-period design) are generally good initial estimates, while the appropriate values of the prior variance have been determined by trial and error. There is scope for further work on the method in order to develop rules of thumb for these, but

this would require work with a large number of data sets, in order to obtain a general view of the situation.

Chapter 10: Discussion

10.1 Introduction

This thesis has been concerned with the problems caused by carry-over in cross-over designs, and has suggested ways of dealing with these problems. The difficulties that carry-over cause are best understood in relation to the simple two-period cross-over design, because this is the simplest such design, and the most widely used and studied. One solution to the problems of carry-over is to use a more complex design rather than this simplistic design, while another is to use a different form of analysis for the design. Both these options have advantages and disadvantages which will be briefly reviewed in this chapter.

10.2 The Problems of Carry-over

Chapter 2 covered the nature of carry-over, and the problems it causes in the analysis of the simple two-period cross-over. The technical details of these problems are explained in more detail in chapter 8, where the power of the test procedure, and the bias in the estimation of the treatment effect are explored. The uncertain nature of carry-over effects contributes to the difficulty in dealing with them. The simplest type of carry-over, in which the active compound in a drug is still having an effect in the next treatment period,

should be very easy to avoid, while any psychological effect of a treatment is at least understandable, but more difficult to counteract. As Freeman[1989] has pointed out, true treatment-period interaction, implying that a treatment has different effects at different times, makes any judgement of the size of a treatment effect impossible, whether a cross-over design is used or not, because the result will, apparently, depend on when the experiment is performed.

10.3 Alternative Designs

Chapter 3 reviewed the considerable literature on the design of cross-over trials, although this thesis has only considered designs for two treatments, because clinical trials most commonly involve such a comparison. Three alternatives to the simple two-period design were dealt with in detail. These are the "complete" two period design, with four sequences AA, BB, AB & BA, shown by Laska et al[1983] to be the optimal two-period design, the three-period designs with two sequences ABB & BAA, shown by the same authors to be the optimal three-period design, and the three-period design with four sequences AAA, BBB, ABB & BAA. This last design does not seem to have been seriously considered before, most authors rejecting the use of a sequence of the same treatment for a cross-over design, despite the optimality of the

"complete" two-period design. This seems unnecessarily restrictive, and the design is worthy of consideration.

All of these alternatives involve some cost, but this is sometimes slight, as it is not necessary to increase the number of subjects in order to obtain a better estimate of the difference between the two treatments than can be obtained from the simple two-period design. With the four sequence designs, this means that there will be a smaller number of subjects in each sequence, which might be a problem if there was a high drop-out rate, although such a phenomenon would be rather alarming anyway. With the three-period designs, there is of course an extra treatment period to administer, suggesting that the trial would last longer which might increase the drop-out rate. However, because these designs cope much better with carry-over, especially first-order carry-over, there is less need for a "wash-out" period between treatments, and so the overall length of the trial may not be much longer than a two-period design with wash-out periods. The " complete" two-period design is the least demanding alternative, only requiring that the subjects be organised into four groups rather than two. Given the benefits in terms of the improved estimation of the treatment difference, this is well worthwhile.

10.4 Alternative Analyses

Much of the thesis is concerned with the statistical analysis of cross-over trials, and how this may be improved. Methods for both binary and continuous data have been considered, with attention being given to classical and Bayesian methods. Almost all analysis options have been considered previously for the simple two-period cross-over, but application to the other designs has often not been discussed before in detail. In particular, results for the ABB, BAA design for binary data is new, (Morrey[1989]), as is the Bayesian analysis using Gibbs sampling. Indeed, the consideration of the efficacy of Gibbs sampling for the analysis of cross-over designs undertaken in chapter 9 has not been performed before. The calculations for the power of the test procedures for the ABB, BAA design contained in chapter 8, have also not been publicised before.

There is considerable scope for further work, both in considering other cross-over designs with, for example, more than two treatments, and other types of data. Although non-parametric methods have not been mentioned in detail in the thesis, most of the tests for continuous data involve a comparison of sets of contrasts from the different sequences, which could be carried out by a non-parametric method as easily as by a parametric method. It would also be possible to modify the Gibbs sampling methods to non-normal likelihoods, although this

would be less trivial. Analysis using multivariate methods has not been considered, although this could be expected to mirror the analysis for continuous data. Perhaps the most difficult area that still needs to be explored in detail is the analysis for ordered categorical data. In general, cross-over designs attempt to extract a great deal of information from relatively few observations. With ordered categorical data, the observations may be regarded as being one stage removed from an underlying, unobservable, latent variable. Thus the observations effectively indicate that the latent variable is in a specific, unknown, range. Given the difficulties of performing the analysis with continuous data, equivalent to observing the latent variable, it is not surprising that analysis is intractable when the latent variable is not directly observable, and it seems doubtful whether a really efficient analysis would ever be possible.

10.5 Cross-over Designs and Clinical Trials

The particular problems of clinical trials arise from the fact that the experimental units are people. The large variability of the experimental units is one consequence, while ethical issues relating to the treatment of subjects cause other problems. It is generally regarded as unethical to treat more subjects than is absolutely necessary with an inferior treatment

in order to provide convincing evidence that would determine good prescriptive practice. Thus a clinical trial should add to the current state of knowledge, rather than duplicating past results, or at least provide substantial weight to a previously tentative conclusion, but in the process it should minimise the number of subjects that receive an inferior treatment. At first sight, the potential gain in precision of a cross-over design would seem to be beneficial, but if the problems of carry-over arise these potential benefits are lost. Indeed, the potential for a misleading estimate of the treatment effect that arises from the suite of pre-tests and main tests commonly used for the simple two-period cross-over is a marked disadvantage. Given the problems with carry-over, the use of the simple two-period cross-over is hard to justify, as the results are unlikely to be completely convincing to any-one familiar with them. This body of people is, however, unlikely to include many physicians, who are more likely to be impressed by the intuitive appeal of the cross-over design. Fortunately, there are cross-over designs, notably the "complete" two-period design, which do not suffer from these disadvantages, and could usefully be employed in clinical trials. Although physicians would need to be informed of the virtues of this design, it could be said to be a combination of the parallel and cross-over designs, combining the benefits of each.

10.6 Concluding Remarks

It is likely that the alluring concept of "using a subject as his own control" will ensure that cross-over designs will continue to be used for clinical trials. It is unfortunate that the appeal of the designs is so intuitively clear, whereas the drawbacks are more difficult to comprehend. This is compounded by the use of the word "carry-over" for the factor that causes all the problems, leading to the impression that such effects can be avoided by an interval between the treatment periods. Most unfortnate of all is the fact that the design which has become most commonly used, the simple two-period cross-over, is badly affected by the problems of carry-over, when there are other, almost equally simple designs which do not suffer from the problem to the same degree. In particular, the "complete" two-period cross-over would be a much better choice for the "standard" two-treatment cross-over design.

REFERENCES

Armitage P.[1955] Tests for linear trends in proportions and frequencies. Biometrics *11* 375-386

Armitage P. & Hills M.[1982] The two-period cross-over trial. The Statistician *31* 119-131

Balaam L.N.[1968] A two-period design with t^2 experimental units. Biometrics *24* 61-73

Berenblut I.I.[1964] Change-over designs with complete balance for first residual effects. Biometrics *20* 707-712

Berenblut I.I.[1967] A change-over design for testing a treatment factor at four equally spaced levels. Jnl. Royal Statist. Soc., B *29* 370-373

Berenblut I.I.[1968] Change-over designs balanced for the linear component of first residual effects. Biometrika *55* 297-303

Bora A.C.[1984] Change-over designs with errors following a first order autoregressive process. Austral. J. Statist. *26* 179-188

Brown B.W.Jr.[1980] The cross-over experiment for clinical trials. Biometrics *36* 69-79

Cochran W.G. Autrey K.M. & Cannon C.Y.[1941] A double change-over design for dairy cattle feeding experiments. Journal of Dairy Science *24* 937-951

Cox D.R.[1970] The analysis of binary data. Chapman & Hall, London

Cox M.A.A. & Plackett R.L.[1980] Matched pairs in factorial experiments with binary data Biometrical Journal *22* 697-702

Daley D.J.[1974] Computation of bi- and tri-variate normal integrals. Applied Statistics *23* 435-438

Donnelly T.G.[1973] Algorithm 462. Bivariate normal distribution. Commun. Assoc. Comput. Mach. *16* 638

Ebbutt A.F.[1984] Three-period cross-over designs for two treatments. Biometrics *40* 219-224

Fava G.M. & Patel H.I.[1986] A survey of cross-over designs used in industry. Unpublished manuscript

Fidler V.[1984] Change-over clinical trials with binary data: mixed-model-based comparison of tests. Biometrics *40* 1063-1070

Fidler V.[1986] Change-over trials with binary data: estimation. Statistica Neerlandica *40* 81-86

Fidler V.[1988] A log-linear model for binary cross-over data. (Letter to the Editors) Applied Statistics *37* 111-112

Freeman P.R.[1989] The performance of the two-stage analysis of two-treatment two-period cross-over trials. Statistics in Medicine *8* 1421-1432

Gart J.J.[1969] An exact test for comparing matched proportions in cross-over designs. Biometrika *56* 75-80

Gelfand A.E. & Smith A.F.M.[1990] Sampling based approaches to calculating marginal densities. JASA *85* 398-409

Gelfand A.E., Hills S.E., Racine-Poon A. & Smith A.F.M.[1990] Illustration of Bayesian inference in normal data models using Gibbs sampling. JASA *85* 972-985

Geman S. & Geman D.[1984] Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence *6* 721-741

Grieve A.P.[1985] A Bayesian analysis of the two-period cross-over design for clinical trials. Biometrica *41* 979-990 (Correction: Biometrics *42* 459)

Grizzle J.E.[1965] The two-period change-over design and its use in clinical trials. Biometrics *21* 467-480 (Correction: Biometrics *30* 727)

Grizzle J.E., Starmer C.F. & Koch G.G.[1969] Analysis of categorical data by linear models. Biometrics *25* 489-504

Hills M. & Armitage P.[1979] The two-period cross-over clinical trial. Brit J. Clin Pharmac *8* 7-20

Huitson A., Poloniecki J., Hews R. & Barker N.[1982] A review of cross-over trials. The Statistician *31* 71-80

Hüsler J. & Lienert G.A.[1985] Treatment evaluation by incremental cross-over designs. Biometrical Journal *27* 47-63

Jones B. & Kenward M.G.[1987] Modelling binary data from a three-period cross-over trial. Statistics in Medicine *6* 555-564

Jones B. & Kenward M.G.[1989] Design and analysis of cross-over trials. Chapman & Hall, London

Keats A.S., Beecher H.K. & Mosteller F.C.[1950] Measurement of pathological pain in distinction to experimental pain. Journal of Applied Physiology *3* 35-44

Kenward M.G. & Jones B.[1987a] A log-linear model for binary cross-over data. Applied Statistics *36* 192-204

Kenward M.G. & Jones B.[1987b] The analysis of data from 2x2 cross-over trials with baseline measurements. Statistics in Medicine *6* 911-926

Kershner R.P. & Federer W.T.[1981] Two treatment cross-over designs for estimating a variety of effects. JASA *76* 612-619

Laska E.E., Meisner M. & Kushner H.B.[1983] Optimal cross-over designs in the presence of carry-over effects. Biometrics *39* 1087-1091

Laska E.E. & Meisner M.[1985] A variational approach to optimal two-treatment cross-over designs: application to carry-over effect models JASA *80* 704-710

Lindley D.V.[1965] Introduction to probability and statistics from a Bayesian viewpoint, part II : inference. Cambridge University Press.

Lindley D.V. & Smith A.F.M.[1972] Bayes estimates for the linear model (with discussion). Jnl. Royal Statist. Soc., B *34* 1-41

Lehmacher W.[1991] Analysis of the cross-over design in the presence of residual effects. Statistics in Medicine *10* 891-899

Lucas H.L.[1957] Extra-period latin-square change-over designs. Journal of Dairy Science *40* 225-239

Mainland D.[1963] Elementary Medical Statistics (2nd edn) Saunders, Philadelphia, Pa.

Matthews J.N.S.[1987] Optimal cross-over designs for the comparison of two treatments in the presence of carry-over effects and autocorrelated errors. Biometrika *74* 311-320

Morrey G.H.[1984] A simple alternative to the two-period cross-over design. Leicester Statistics Technical Report no. 2

Morrey G.H.[1989] Binary responses and the three-period cross-over Biometrical Journal *31* 589-598

Naylor J.C. & Smith A.F.M.[1982] Applications of a method for the efficient computing of posterior distributions. Applied Statistics *31* 214-225

Naylor J.C. & Smith A.F.M.[1988] Economic illustrations of novel numerical integration strategies for Bayesian inference. Journal of Econometrics *38* 103-126

Nelder J.A.[1977] A reformulation of linear models. (with discussion) Jnl. Royal Statist. Soc. A *140* 48-76

Nelder J.A.[1982] Linear models and non-orthogonal data. Utilitas Mathematica *21* 141-152

Owen D.B.[1956] Tables for computing bivariate normal probabilities. Ann. Math. Statist. *27* 1075-1090

Patterson H.D.[1952] The construction of balanced designs for experiments involving sequences of treatments. Biometrika *39* 32-48

Patterson H.D. & Lucas H.L.[1959] Extra-period change-over designs. Biometrics *15* 116-132

Pendleton O.J., Von Tress M. & Bremer R.[1986] Interpretation of the four types of variance tables in SAS. Commun. Statist. -Theor. Meth. *15* 2785-2808

Plackett R.L.[1974] The analysis of categorical data. Griffin, London

Poloniecki J. & Daniel D.[1981] Further analysis of Hills and Armitage enuresis data. The Statistician *30* 225-229

Prescott R.J.[1981] The comparison of success rates in cross-over trials in the presence of an order effect. Applied Statistics *30* 9-15

Racine A., Grieve A.P., Flühler H. & Smith A.F.M.[1986] Bayesian methods in practice: experience in the pharmacuetical industry (with discussion). Applied Statistics *35* 93-150

Selwyn M.R., Dempster A.P. & Hall N.R.[1981] A Bayesian approach to bioequivalence for the 2x2 change-over design. Biometrics *37* 11-21

Smith A.F.M.[1973] A general Bayesian linear Model. Jnl. Royal Statist. Soc., B *35* 67-75

Steck G.P.[1958] A table for computing trivariate normal probabilities. Ann. Math. Statist. *29* 780-800

Williams E.J.[1949] Experimental designs balanced for the estimation of residual effects of treatments. Aust. J. Scient. Res.(A) *2* 149-168

Williams E.J.[1950] Experimental designs balanced for pairs of residual effects. Aust. J. Scient. Res.(A) *3* 351-363

Zimmerman H. & Rahlfs V.[1978] Testing hypotheses in the two-period change-over with binary data. Biometrical Journal *20* 133-141

APPENDICES

Appendix A Calculation of Multivariate Normal Probabilities

Appendix B Program to Calculate the Power for the Two-period Cross-over

Appendix C Program to Calculate the Power for the Three-period Cross-over

Appendix D Derivation of Posterior Distributions for the Two-period Cross-over

Appendix E Derivation of Posterior Distributions for the Three-period Two-sequence Cross-over

Appendix F Data Sets

APPENDIX A

Calculation of Multivariate Normal Probabilities

Appendix A : Calculation of Multivariate Normal Probabilities

Introduction

In order to calculate the power of the test procedures for cross-over designs where a sequence of pre-tests and main tests are employed, calculation of bivariate and tri-variate normal probabilities is required. The purpose of this appendix is to describe the methods that were used to perform these calculations.

Bivariate Normal Probabilities

Various methods of calculating bivariate normal probabilities have been described, some of which apply in special circumstances, but one of the most useful methods is given by Owen[1956], and has been used in this study. Defining $B(h,k;\rho)$ to be the bivariate normal integral for the volume $X < h$, $Y < k$ when variates X and Y have a bivariate normal distribution with $E(X) = E(Y) = 0$, $Var(X) = Var(Y) = 1$, and $Cov(XY) = \rho$, and $G(z)$ to be the probability that the standard normal variate Z takes a value $< z$, Owen showed that:

$B(h,k;\rho) = \frac{1}{2}G(h) + \frac{1}{2}G(k) - T(h,a_h) - T(k,a_k) - b$

where the constant b is 0 if $hk > 0$ or if $hk = 0$ but $h+k \geq 0$, and $\frac{1}{2}$ otherwise, and the function $T(h,a)$ is an integral defined by Owen, with $a_h = ((k/h) - \rho)/r$,

$a_k = ((h/k) - \rho)/r$ where $r = \sqrt{(1 - \rho^2)}$. Owen also gave a series expansion for the integral $T(h,a)$, which rapidly converges for small values of a and h. Donnelly[1973] has given a Fortran program for calculating bivariate normal probabilities which evaluates this infinite series, and a modified version of this has been used for the calculations, this forms the subroutine "BCAL" in the programs to calculate the power. A standard NAG function (S15ABF) was used to evaluate standard normal probabilities.

Trivariate Normal Probabilities

In order to calculate trivariate normal probabilities, a method described by Steck[1958] was applied. If X, Y & Z are three variables having a trivariate normal distribution with $E(X) = E(Y) = E(Z) = 0$, $Var(X) = Var(Y) = Var(Z) = 1$, and $E(XY) = \rho_{12}$, $E(XZ) = \rho_{13}$, $E(YZ) = \rho_{23}$; the probability that $X < h$, $Y < k$, and $Z < m$ is $C(h,k,m;\rho_{12},\rho_{13},\rho_{23})$. This probability is defined in terms of the G and T functions given above, and a new function, $S(h,a,b)$, which is itself a univariate integral. Daley[1974] has commented that computation of this integral by Gaussian quadrature gives acceptably small errors, so a NAG routine (D01ARF) was used to perform the evaluation of the S function, and hence to obtain trivariate normal probabilities. Calculation of the S function is performed in the subroutine "SCAL" in

the program to calculate the power for the three-period cross-over design.

$C(h,k,m;\rho a_{12},\rho_{13},\rho_{23})$ is only defined for two cases. Case 1 applies if h,k, & m are all positive or all negative, and case 2 applies if h & k are positve and m negative or h & k negative and m positive. It is possible to interchange the variables, and the corresponding correlations, in order to ensure that it is always m that has a different sign. Because $C(h,k,m)$ is defined as $p(X<h,\ Y<k,\ Z<m)$, some re-adjustment needs to be made if a probability such as $p(X<h,\ Y<k,\ Z>m)$ is required. Expressions for the required probability can be found by using the fact that, because $C(h,k,m)$ applies to normalised variables X, Y & Z, $C(h,k,m;\rho_{12},\rho_{13},\rho_{23}) = D(-h,-k,-m;\rho_{12},\rho_{13},\rho_{23})$, where $D(h,k,m) = p(X>h,\ Y>k,\ Z>m)$, and the relationship given by Steck that if h,k,m are all non-negative or non-positive. $c(h,k,m;\rho_{12},\rho_{13},\rho_{23}) = B(h,k;\rho_{12}) - C(h,k,-m;\rho_{12},-\rho_{13},-\rho_{23})$.

APPENDIX B

Program to Calculate the Power for the Two-period Cross-over

Appendix B : Program to calculate the power for the two-period cross-over.

```
c programme to calculate the power for the 2-period x-over
c
      integer ifail,ilp,jlp
      double precision n,rn,r12,r32,r,d,g,v1,v2,v3,v4,v5,v6,h,k,pow,
     1biv,t,x,s15abf,tm1,tm2,res(17,17),that,xinc
      common ifail,biv,t
      read(5,10) n,za1,za2,za3
   10 format(4d15.8)
   12 rn=sqrt(n)
      r12=sqrt(1.2d1)
      r32=sqrt(3.0d0)/2.0d0
      r=r32
      d=1.0d0
      g=-4.2d0
      do 30 ilp = 1,41
      g=g+0.2d0
      pow=1.0d0
      v2=za2-rn*d/2
      v3=za2+rn*d/2
      v1=za1-rn*g/r12
      v4=za1+rn*g/r12
      v5=za3-(d-g/2)*rn
      v6=-za3-(d-g/2)*rn
      x=v1
      ifail=0
      p=s15abf(x,ifail)
      pow=pow-p
      tm1=p
      x=v2
      ifail=0
      p=s15abf(x,ifail)
      pow=pow-p
      call bcal(v1,v2,r)
      pow=pow+biv
      x=-v3
      ifail=0
      p=s15abf(x,ifail)
      pow=pow+p
      call bcal(v1,-v3,r)
      pow=pow-biv
      x=-v4
      ifail=0
      p=s15abf(x,ifail)
      pow=pow+p
      call bcal(-v4,v2,r)
      pow=pow-biv
      call bcal(-v4,-v3,r)
      pow=pow+biv
      x=-v4
      ifail=0
      p=s15abf(x,ifail)
      tm1=tm1-p
c
c                B-1
c
```

```
c
c
c
c
c
c
c
c
c
      x=v5
      ifail=0
      p=s15abf(x,ifail)
      tm2=1-p
      x=v6
      ifail=0
      p=s15abf(x,ifail)
      tm2=tm2+p
      pow=pow+tm1*tm2
      that=tm1*(d-g/2) + (1-tm1)*d-0.5d0
      write(6,29) g,tm1,that,pow
   29 format(5d15.8)
   30 continue
      stop
      do 40 ilp=1,17
      do 40 jlp=1,17
      write(6,50) res(ilp,jlp)
   40 continue
   45 format(13htreat diff = ,d15.8)
   50 format(d15.8)
      end
c
c  subroutine to calculate bivariate normals
c
      subroutine bcal(bh2,bk2,br2)
      integer ifail
      double precision bh,bk,br,qr,biv,p,ah,ak,t,s15abf,x,bh2,bk2,br2
      common ifail,biv,t
      biv=0.0d0
      bh=bh2
      bk=bk2
      br=br2
      qr=sqrt(1-br*br)
      x=bh
      ifail=0
      p=s15abf(x,ifail)
      biv=biv+p
      x=bk
      ifail=0
      p=s15abf(x,ifail)
      biv=biv+p
      biv=biv/2.0d0
      ah=(bk/bh-br)/qr
      call tcal(bh,ah)
      biv=biv-t
      ak=(bh/bk-br)/qr
      call tcal(bk,ak)
      biv=biv-t
      if (bh*bk) 110,100,120
  100 if (bh+bk) 110,120,120
  110 biv=biv-0.5d0
  120 return
      end
c
c
c
```

```
c
c
c
c
c
c
c
c
c
c
c
c
c
c  subroutine to calculate t-function
c
      subroutine tcal(th2,ta2)
      integer ifail,flg
      double precision th,ta,twopi,con,gw,t,sgn,g2,h2,aq,h4,ex,w2,ap,
     1s2,sp,s1,sn,conex,cn,x,p,s15abf,b,th2,ta2,biv
      common ifail,biv,t
      th=th2
      ta=ta2
      twopi=6.283185307179587d0
      con=twopi*1d-15/2.0d0
      b=0.0d0
      flg=0
      s1=0.0d0
      sgn=-1.0d0
      if (th) 190,300,200
  190 th=-th
  200 x=th
      ifail=0
      p=s15abf(x,ifail)
      gw=p
      wh=th
      t=0.0d0
      if (ta) 210,320,220
  210 ta=-ta
      flg=1
  220 if (dabs(ta)-1.0d0) 270,230,240
  230 t=ta*gw*(1.0d0-gw)/2.0d0
      goto 310
  240 sgn=-sgn
      wh=wh*ta
      x=wh
      p = s15abf(x,ifail)
      g2=p
      ta=1.0d0/ta
  260 b=b+(gw+g2)/2.0d0-gw*g2
  270 h2=wh*wh
      aq=ta*ta
      h4=h2/2.0d0
      ex=dexp(-h4)
      w2=h4*ex
      ap=1.0d0
      s2=ap-ex
      sp=ap
      sn=s1
      conex=dabs(con/ta)
      goto 290
  280 sn=sp
      sp=sp+1.0d0
      s2=s2-w2
      w2=w2*h4/sp
c
c                B-3
c
c
```

```
      ap=-ap*aq
  290 cn=ap*s2/(sn+sp)
      s1=s1+cn
      if (dabs(cn)-conex) 300,300,280
  300 t=(datan(ta)-ta*s1)/twopi
  310 t=b-sgn*t
      if (flg) 330,330,315
  315 t=-t
      goto 330
  320 t=0.0d0
  330 return
      end
```

APPENDIX C

Program to Calculate the Power for the Three-period Cross-over

c
c
c
c
c
c
c
c
c
c
c
c
c
c

Appendix C : Program to calculate the power for the three-period cross-over

```
c   programme to calculate the power for the three=period cross-over
c
      integer ifail,i
      double precision ans,biv,cval,n,p,pow,pr1,pr2,pr3,
     1q44,ret,rq33,rq44,rtt,rn,rn3,t,tre,v1,v2,v3,v4,v5,
     2v6,v7,v8,v9,v10,s15abf,x,zer,d,f,g,rrn,pk1,pk2,pk3,
     3dhat,tem
      common/block1/ans,biv,cval,t
      ret=dsqrt(11.0d0/12.0d0)
      rtt= dsqrt(3.0d0)/2.0d0
      rq33=1.0d0/dsqrt(33.0d0)
      tre=3.0d0/dsqrt(11.0d0)
      rq44=1.0d0/dsqrt(44.0d0)
      zer=0.0d0
      read(5,10)n,za1,za2,za3,za4,za5
   10 format(6d10.4)
  411 format(2d10.4)
  412 format(8h cry1 = ,d10.4,8h cry2 = ,d10.4)
      write(6,*) n
      write(6,*) za1,za2,za3,za4,za5
      write(6,412) g,f
c
c  d=treat diff, f=carry 2 diff, g=carry 1 diff
c
      rn=dsqrt(n)
      rn3=dsqrt(n/3.0d0)
      rrn=1.0d0/rn
      f=5.0d0*rrn
      g=-5.0d0*rrn
      write(6,*) g,f
      d=-6.2d0*rrn
      do 8 i=1,61
      d=d+rrn/5.0d0
      v1=-za1-f*rn*rq44
      v2=za1-f*rn*rq44
      v3=-za2-g*rn3/2.0d0
      v4=za2-g*rn3/2.0d0
      v5=-za3-d*rn/2.0d0
      v6=za3-d*rn/2.0d0
      v7=-za4-(d-g/2.0d0)*rn
      v8=za4-(d-g/2.0d0)*rn
      v9=-za5-(d-f/4.0d0)*2.0d0*rn3
      v10=za5-(d-f/4.0d0)*2.0d0*rn3
      pow=0.0d0
c
c  calculation of prob U not sig, V sig
c
      x=v2
      ifail=0
      p=s15abf(x,ifail)
      pr1=p
      x=v1
c
c
c
```

```
c
c
c
c
c
c
c
c
c
c
c
c
      ifail=0
      p=s15abf(x,ifail)
      pr1=pr1-p
      pk1=pr1
      call bcal(v2,v10,rq33)
      pr1=pr1-ans
      call bcal(v2,v9,rq33)
      pr1=pr1+ans
      call bcal(v1,v10,rq33)
      pr1=pr1+ans
      call bcal(v1,v9,rq33)
      pr1=pr1-ans
c
c  calculation of prob U sig, R not sig
c
      x=v4
      ifail=0
      p=s15abf(x,ifail)
      pk2=p
      x=v3
      ifail=0
      p=s15abf(x,ifail)
      pk2=pk2-p
      call bcal(v2,v4,ret)
      pk2=pk2-ans
      call bcal(v1,v4,ret)
      pk2=pk2+ans
      call bcal(v2,v3,ret)
      pk2=pk2+ans
      call bcal(v1,v3,ret)
      pk2=pk2-ans
c
c  calculation of prob U sig, R sig and E(delta)
c
      pk3=1.0d0-pk1-pk2
      dhat=pk1*(d-f/4.0d0)+pk2*(d-g/2.0d0)+pk3*d
c
c  calculation of prob U sig, R sig, S sig
c
      call trical(v1,v3,v5,ret,tre,rtt,0,1)
      pr2=cval
      call trical(v1,v3,v6,ret,tre,rtt,0,4)
      pr2=pr2+cval
      call trical(v1,v4,v5,ret,tre,rtt,0,3)
      pr2=pr2+cval
      call trical(v1,v4,v6,ret,tre,rtt,1,2)
      pr2=pr2+cval
      call trical(v2,v3,v5,ret,tre,rtt,0,2)
      pr2=pr2+cval
      call trical(v2,v3,v6,ret,tre,rtt,1,3)
      pr2=pr2+cval
      call trical(v2,v4,v5,ret,tre,rtt,1,4)
c
c                C-2
c
c
```

```
c
c
c
c
c
c
c
c
c
c
c
      pr2=pr2+cval
      call trical(v2,v4,v6,ret,tre,rtt,1,1)
      pr2=pr2+cval
c
c  calculation of prob U sig, R not sig, T sig
c
      call trical(v1,v4,v7,ret,rq44,zer,0,1)
      pr3=cval
      call trical(v1,v3,v7,ret,rq44,zer,0,1)
      pr3=pr3-cval
      call trical(v1,v4,v8,ret,rq44,zer,0,4)
      pr3=pr3+cval
      call trical(v1,v3,v8,ret,rq44,zer,0.4)
      pr3=pr3-cval
      call trical(v2,v4,v7,ret,rq44,zer,0,2)
      pr3=pr3+cval
      call trical(v2,v3,v7,ret,rq44,zer,0,2)
      pr3=pr3-cval
      call trical(v2,v4,v8,ret,rq44,zer,1,3)
      pr3=pr3+cval
      call trical(v2,v3,v8,ret,rq44,zer,1,3)
      pr3=pr3-cval
      pow=pr1+pr2+pr3
      if (i.ne.1) goto 420
  420 continue
      dhat=dhat-0.5d0
      if (pow.lt.1.0d0) goto 425
      pow=1.0d0
  425 continue
      tem=d/rrn
      write(6,430) tem,pk2,dhat,pow
  430 format(4d15.8)
   11 format(/7h power ,d15.8,3h ; ,3(d15.8,4x))
   12 format(11h delta est ,4(d15.8,4x))
    9 format(5h d = ,d15.8,5h f = ,d15.8,5h g = ,d15.8)
    8 continue
      stop
      end
c
c  subroutine to calculate tri-variate normals
c
      subroutine trical(ht,kt,mt,rt12,rt13,rt23,ia,ib)
      integer ifail,ind,ia,ib,id,ipl
      double precision a1,a2,a3,b1,b2,b3,c1,c2,c3,d1,d2,d3,f1,f2,f3,
     1e1,e2,e3,e4,e5,e6,h,k,m,r12,r13,r23,rc1,rc2,rc3,biv,del,gt,tt,st,
     2tri,ans,t,x,xx1,xx2,cval,p,s15abf,ht,kt,mt,rt12,rt13,rt23
      common/block1/ans,biv,cval,t
      common/block2/h,k,m,r12,r13,r23
      tri=0.0d0
      id=0
c
c                C-3
c
c
c
c
c
c
```

```
c
c
c
c
c
c
c
c  ia=0 if more < than >  ia=1 if more > than <
c
      h=ht
      k=kt
      m=mt
      r12=rt12
      r13=rt13
      r23=rt23
      if (ia.eq.0) goto 12
      h=-h
      k=-k
      m=-m
      ipl=1
c
c  now more < than >. ib=1 for all < ib=2 for x> ib=3 for y> ib=4 for z>
c
   12 goto (15,13,14,15),ib
   13 ipl=2
      call sw13
      goto 15
   14 ipl=3
      call sw23
c
c  now all < or only z>
c
   15 ind=1
      if (h.lt.0) ind=ind+1
      if (k.lt.0) ind=ind+2
      if (m.lt.0) ind=ind+3
      if (ind.gt.4) ind=8-ind
      ipl=4
c
c  ind=1 all limits same sign, ind=2 h is odd, ind=3 k is odd, ind=4 m is odd
c
      if (ib.eq.1) goto 30
      goto (21,20,20,22),ind
   20 id=1
      ipl=5
   21 call bcal(h,k,r12)
      tri=biv
      ipl=6
      goto 30
   22 call cs3
      ind=1
      ipl=7
   30 goto (40,32,33,34),ind
   32 call sw13
      ipl=8
      goto 34
   33 call sw23
      ipl=9
      goto 34
c
c                C-4
c
c
c
c
c
c
```

```
c
c
c
c
c
c
   34 call bcal(h,k,r12)
      ipl=9
      if (id.eq.1) goto 35
      tri=biv
      goto 36
   35 tri=tri-biv
      ipl=10
   36 call cs3
      ipl=11
   40 xx1=1.0d0
      xx2=1.0d0
      e1=k-h*r12
      e2=m-k*r23
      e3=h-m*r13
      e4=m-h*r13
      e5=h-k*r12
      e6=k-m*r23
      f1=1.0d0-r12*r12
      f2=1.0d0-r23*r23
      f3=1.0d0-r13*r13
      rc1=r12-r13*r23
      rc2=r13-r12*r23
      rc3=r23-r12*r13
      a1=e1/(h*sqrt(f1))
      a2=e2/(k*sqrt(f2))
      a3=e3/(m*sqrt(f3))
      c1=e4/(h*sqrt(f3))
      c2=e5/(k*sqrt(f1))
      c3=e6/(m*sqrt(f2))
      del=sqrt(f2-r12*rc1-r13*rc2)
      b1=(f1*e4-rc3*e1)/(e1*del)
      d1=(f3*e1-rc3*e4)/(e4*del)
      b2=(f2*e5-rc2*e2)/(e2*del)
      d2=(f1*e2-rc2*e5)/(e5*del)
      b3=(f3*e6-rc1*e3)/(e3*del)
      d3=(f2*e3-rc1*e6)/(e6*del)
c
c  gt=g-term
c
   50 gt=0.0d0
      xx1=dsign(xx1,a1)
      xx2=dsign(xx2,c1)
      if ((xx1*xx2).ne.1.0d0) goto 100
      x=h
      ifail=0
      p=s15abf(x,ifail)
      gt=gt+p
  100 xx1=dsign(xx1,a2)
      xx2=dsign(xx2,c2)
      if ((xx1*xx2).ne.1.0d0) goto 105
      x=k
      ifail=0
c
c                C-5
c
c
c
c
c
c
c
```

```
c
c
c
c
c
      p=s15abf(x,ifail)
      gt=gt+p
  105 xx1=dsign(xx1,a3)
      xx2=dsign(xx2,c3)
      if((xx1*xx2).ne.1.0d0) goto 110
      x=m
      ifail=0
      p=s15abf(x,ifail)
      gt=gt+p
  110 gt=gt/2.0d0
c
c  tt=t-term
c
      tt=0.0d0
      call tcal(h,a1)
      tt=tt+t
      call tcal(h,c1)
      tt=tt+t
      call tcal(k,a2)
      tt=tt+t
      call tcal(k,c2)
      tt=tt+t
      call tcal(m,a3)
      tt=tt+t
      call tcal(m,c3)
      tt=tt+t
c
c  st=s-term
c
      st=0.0d0
      call scal(h,a1,b1)
      st=st+ans
      call scal(h,c1,d1)
      st=st+ans
      call scal(k,a2,b2)
      st=st+ans
      call scal(k,c2,d2)
      st=st+ans
      call scal(m,a3,b3)
      st=st+ans
      call scal(m,c3,d3)
      st=st+ans
      cval=gt-tt/2.0d0-st
      ipl=20
      if (id.eq.1) goto 120
      if (tri.ne.0) cval=tri-cval
      goto 130
  120 cval=tri+cval
  130 return
      end
c
c                C-6
c
c
c
c
c
c
c
c
c
c
c
```

```
c
c
      ifail=0
      p=s15abf(x,ifail)
      biv=biv+p
      x=bk
      ifail=0
      p=s15abf(x,ifail)
      biv=biv+p
      biv=biv/2.0d0
      ah=(bk/bh-br)/qr
      call tcal(bh,ah)
      biv=biv-t
      ak=(bh/bk-br)/qr
      call tcal(bk,ak)
      biv=biv-t
      if (bh*bk) 110,100,120
  100 if (bh+bk) 110,120,120
  110 biv=biv-0.5d0
  120 ans=biv
      return
      end
c
c  subroutine to calculate Owen's t-function
c
      subroutine tcal(th2,ta2)
      integer ifail,flg
      double precision th,ta,twopi,con,gw,t,tt,sgn,g2,h2,aq,h4,ex,w2,ap,
     1s2,sp,s1,sn,conex,cn,x,p,s15abf,b,ans,th2,ta2,biv,wh,cval
      common/block1/ans,biv,cval,t
      th=th2
      ta=ta2
      twopi=6.2831853071795867d0
      con=twopi*1d-15/2.0d0
      b=0.0d0
      flg=0
      s1=0.0d0
      sgn=-1.0d0
      if (th) 190,300,200
  190 th=-th
  200 x=th
      ifail=0
      p = s15abf(x,ifail)
      gw=p
      wh=th
      t=0.0d0
      if (ta) 210,320,220
  210 ta=-ta
      flg=1
  220 if (dabs(ta)-1.0d0) 270,230,240
  230 t=ta*gw*(1.0d0-gw)/2.0d0
      goto 310
  240 sgn=-sgn
      wh=wh*ta
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
```

```
c
c
c
c  subroutine to swap x & z
c
      subroutine sw13
      double precision h,k,m,r12,r13,r23,tms
      common/block2/h,k,m,r12,r13,r23
      tms=h
      h=m
      m=tms
      tms=r12
      r12=r23
      r23=tms
      return
      end
c
c  subroutine to swap y & z
c
      subroutine sw23
      double precision h,k,m,r12,r13,r23,tms
      common/block2/h,k,m,r12,r13,r23
      tms=k
      k=m
      m=tms
      tms=r12
      r12=r13
      r13=tms
      return
      end
c
c  subroutine to change sign of z and r13,r23
c
      subroutine cs3
      double precision h,k,m,r12,r13,r23
      common/block2/h,k,m,r12,r13,r23
      m=-m
      r13=-r13
      r23=-r23
      return
      end
c
c  subroutine to calculate bivariate normals
c
      subroutine bcal(bh2,bk2,br2)
      integer ifail
      double precision bh,bk,br,qr,biv,p,ah,ak,t,s15abf,x,
     1bh2,bk2,br2,ans,cval
      common/block1/ans,biv,cval,t
      biv=0.0d0
      bh=bh2
      bk=bk2
      br=br2
      qr=dsqrt(1-br*br)
      x=bh
c
c                C-8
c
c
c
c
c
c
c
c
c
c
c
```

```
c
      x=wh
      p = s15abf(x,ifail)
      g2=p
      ta=1.0d0/ta
  260 b=b+(gw+g2)/2.0d0-gw*g2
  270 h2=wh*wh
      aq=ta*ta
      h4=h2/2.0d0
      ex=dexp(-h4)
      w2=h4*ex
      ap=1.0d0
      s2=ap-ex
      sp=ap
      sn=s1
      conex=dabs(con/ta)
      goto 290
  280 sn=sp
      sp=sp+1.0d0
      s2=s2-w2
      w2=w2*h4/sp
      ap=-ap*aq
  290 cn=ap*s2/(sn+sp)
      s1=s1+cn
      if (dabs(cn)-conex) 300,300,280
  300 t=(datan(ta)-ta*s1)/twopi
  310 t=b-sgn*t
      if (flg) 330,330,315
  315 t=-t
      goto 330
  320 t=0.0d0
  330 return
      end
c
c  subroutine to calculate Stack's s-function
c
      subroutine scal(sh,sg,b)
      double precision a,absac,acc,b,relacc,g,hh,fun,s15abf,alpha(390),
     1t,tt,st,ans,sh,sg,biv,cval
      external fun
      integer iparm,n,ifail
      common/block1/ans,biv,cval,t
      common/block3/g,hh
      relacc=0.0d0
      absac=1.0d-5
      iparm=0
      ifail=1
      a=0.0d0
      hh=sh
      g=sg
      call d01arf(a,b,fun,relacc,absac,0,iparm,acc,ans,n,
     1alpha,ifail)
      if (ifail.eq.0) goto 400
c
c                C-9
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
```

```
      write(6,999) ifail
      stop
  400 return
  999 format(22h d01arf fails. ifail = ,i2)
      end
c
c  routine to calculate values for the s-function
c
      double precision function fun(x)
      double precision x,ta,z,q1,g,hh,s15abf
      common/block3/g,hh
      integer ifail
      ta=dsqrt(1+g*g*(1+x*x))
      z=hh*ta
      ifail = 0
      q1 = s15abf(z,ifail)
      fun=q1/(2*3.141593*ta*(1+x*x))
      return
      end
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
c
```

APPENDIX D

Derivation of Posterior Distributions for the Two-period Cross-over

Appendix D : Derivation of Posterior Distributions for the Two-period Cross-over

Posterior Distribution of θ

Results given in Lindley & Smith[1972] and Smith[1973] state that if $\mathbf{Y}$ is a vector of observations and θ a vector of unknown parameters such that $\mathbf{Y} \sim N(\mathbf{X}\theta, \mathbf{S})$, where $\mathbf{X}$ is a design matrix, and $\mathbf{S}$ is the variance-covariance matrix of the observations, with θ having a prior distribution specified by $\theta \sim N(\mathbf{M}, \mathbf{C})$, then the posterior distribution of θ, given $\mathbf{X}$, $\mathbf{S}$, $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{Y}$ is $N(\theta^*, \mathbf{D})$ with

$$\theta^* = \mathbf{D}(\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y} + \mathbf{C}^{-1}\mathbf{M}) \text{ and } \mathbf{D}^{-1} = \mathbf{X}'\mathbf{S}^{-1}\mathbf{X} + \mathbf{C}^{-1}$$

Since the observations on different subjects are independent, the variance-covariance matrix of the observations, $\mathbf{S}$ takes the form $\mathbf{S} = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \Sigma \end{pmatrix}$

where Σ is a pxp variance-covariance matrix for the p observations on a single subject, and 0 is a pxp matrix of zeros. Similarly, the design matrix $\mathbf{X}$ will consist of n_1 sub-matrices $\mathbf{X}_1$ followed by n_2 submatrices $\mathbf{X}_2$ for the observations for subjects in sequences 1 and 2 respectively. Hence $\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y} = \Sigma\mathbf{X}_i'\Sigma^{-1}\mathbf{Y}_i$, and $\mathbf{X}'\mathbf{S}^{-1}\mathbf{X} = \Sigma\mathbf{X}_i'\Sigma^{-1}\mathbf{X}_i$

For the two-period cross-over without carry-over $\theta' = (\mu, \pi, \tau)$, with $\pi_1 = -\pi_2 = \pi$, and $\tau_A = -\tau_B = \tau$, and

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \qquad \mathbf{X}_2 = \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

and $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_1^2 \end{pmatrix}$ so that $\Sigma^{-1} = (1/\sigma_1^2\sigma_2^2) \begin{pmatrix} \sigma_1^2 & -\sigma_0^2 \\ -\sigma_0^2 & \sigma_1^2 \end{pmatrix}$

where $\sigma_1^2 = \sigma_0^2 + \sigma_e^2$ and $\sigma_2^2 = 2\sigma_0^2 + \sigma_e^2$

Hence, for a subject in sequence 1:

$$\mathbf{X}'\Sigma^{-1}\mathbf{Y} = \begin{pmatrix} (y_{ij1} + y_{ij2})/\sigma_2^2 \\ (y_{ij1} - y_{ij2})/\sigma_e^2 \\ (y_{ij1} - y_{ij2})/\sigma_e^2 \end{pmatrix} \qquad \mathbf{X}'\Sigma^{-1}\mathbf{X} = 1/(\sigma_e^2\sigma_2^2) \begin{pmatrix} 2\sigma_e^2 & 0 & 0 \\ 0 & 2\sigma_1^2 & 2\sigma_1^2 \\ 0 & 2\sigma_1^2 & 2\sigma_1^2 \end{pmatrix}$$

and for a subject in sequence 2:

$$\mathbf{X}'\Sigma^{-1}\mathbf{Y} = \begin{pmatrix} (y_{ij1} + y_{ij2})/\sigma_2^2 \\ (y_{ij1} - y_{ij2})/\sigma_e^2 \\ (-y_{ij1} + y_{ij2})/\sigma_e^2 \end{pmatrix} \qquad \mathbf{X}'\Sigma^{-1}\mathbf{X} = 1/(\sigma_e^2\sigma_2^2) \begin{pmatrix} 2\sigma_e^2 & 0 & 0 \\ 0 & 2\sigma_1^2 & -2\sigma_1^2 \\ 0 & -2\sigma_1^2 & 2\sigma_1^2 \end{pmatrix}$$

so that $\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y} = \begin{pmatrix} 2\Sigma\Sigma (y_{ij1} + y_{ij2})/\sigma_2^2 \\ 2\Sigma\Sigma (y_{ij1} - y_{ij2})/\sigma_e^2 \\ 2\{\Sigma (y_{ij1} - y_{ij2}) - \Sigma (y_{2j1} - y_{2j2})\}/\sigma_e^2 \end{pmatrix}$

and $\mathbf{X}'\mathbf{S}^{-1}\mathbf{X} = \begin{pmatrix} 2(n_1 + n_2)/\sigma_2^2 & 0 & 0 \\ 0 & 2(n_1 + n_2)/\sigma_e^2 & 2(n_1 - n_2)/\sigma_e^2 \\ 0 & 2(n_1 - n_2)/\sigma_e^2 & 2(n_1 + n_2)/\sigma_e^2 \end{pmatrix}$

This allows the calculation of $\mathbf{D}^{-1}$, and hence $\mathbf{D}$, by the use of a NAG matrix inversion routine, (f01aaf or f01acf) so that the mean vector θ^* of the posterior distribution for θ can be calculated. Values from this posterior distribution are simulated using the NAG routines g05eaf and g05ezf.

Posterior Distributions for σ_ε^2 & σ_δ^2

New variables, $y_{\bar{1}j}$ and $y_{\hat{1}j}$ are defined as follows:

$$y_{\bar{1}j} = \tfrac{1}{2}(y_{1j1} - y_{1j2})$$

$$y_{\hat{1}j} = \tfrac{1}{2}(y_{1j1} + y_{1j2})$$

These new variables have the advantage of being orthogonal, so that the bivariate normal distribution of $y_{\bar{1}j}$ and $y_{\hat{1}j}$ is relatively simple. Thus

$$\begin{pmatrix} y_{\bar{1}j} \\ y_{\hat{1}j} \end{pmatrix} \sim N \begin{pmatrix} \pi+\tau \\ \mu \end{pmatrix} \tfrac{1}{2} \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} y_{\bar{2}j} \\ y_{\hat{2}j} \end{pmatrix} \sim N \begin{pmatrix} \pi-\tau \\ \mu \end{pmatrix} \tfrac{1}{2} \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

where $\sigma_2^2 = 2\sigma_\delta^2 + \sigma_\varepsilon^2$

Using Σ_{AB} and Σ_{BA} to denote the summation over subjects in the AB & BA sequences, respectively, the likelihood for the 2N observations on the N subjects can be written in terms of these new variables to give :

$$\pi^{-N}\, \sigma_\varepsilon^{-N}\, \sigma_2^{-N}\, \exp[-\Sigma_{AB}\{y_{\bar{1}j} - (\pi+\tau)\}^2/\sigma_\varepsilon^2] * \exp[-\Sigma_{AB}\{y_{\hat{1}j} - \mu\}^2/\sigma_2^2]$$
$$* \exp[-\Sigma_{BA}\{y_{\bar{1}j} - (\pi-\tau)\}^2/\sigma_\varepsilon^2] * \exp[-\Sigma_{BA}\{y_{\hat{1}j} - \mu\}^2/\sigma_2^2]$$

Given that the prior distribution of θ, σ_ε^2, and σ_2^2 is given by:

$$f(\theta, \sigma_\varepsilon^2, \sigma_2^2) = N(M, C)\, IG(\tfrac{1}{2}d_\varepsilon, \tfrac{1}{2}d_\varepsilon v_\varepsilon)\, IG(\tfrac{1}{2}d_2, \tfrac{1}{2}d_2 v_2)$$

where $IG(\frac{1}{2}d, \frac{1}{2}dv)$ denotes the inverse gamma distribution with parameters $\frac{1}{2}d$ and $\frac{1}{2}dv$, which has a pdf given by:

$$p(\sigma^2) \propto (\sigma^2)^{-(\frac{1}{2}d+1)} \exp[-(dv)/(2\sigma^2)]$$

The joint posterior distribution is proportional to the product of the prior and likelihood given above. The conditional posterior distribution of one of the parameters, given the values of the others, can be found by treating all these other parameters as fixed in the expression for the joint posterior. Thus, the posterior distribution for σ_ε^2, given the values of σ_α^2, θ, and the observations Y can be obtained by picking out the terms involving σ_ε^2, and ignoring those that do not. This gives:

$$(\sigma_\varepsilon^2 | \mathbf{Y}, \theta, \sigma_\alpha^2) \propto (\sigma_\varepsilon^2)^{-(N d_\varepsilon + N/2 + 1)}$$
$$* \exp[-(d_\varepsilon v_\varepsilon + 2\Sigma_{AB}\{y_{ij} - (\pi+\tau)\}^2 + 2\Sigma_{BA}\{y_{ij} - (\pi-\tau)\}^2)/2\sigma_\varepsilon^2]$$

Defining $SS_1 = 2\Sigma_{AB}\{y_{ij} - (\pi+\tau)\}^2 + 2\Sigma_{BA}\{y_{ij} - (\pi-\tau)\}^2$, this may be written:

$$(\sigma_\varepsilon^2 | \mathbf{Y}, \theta, \sigma_\alpha^2) \propto (\sigma_\varepsilon^2)^{-(N(d_\varepsilon+N)+1)} * \exp[-(d_\varepsilon v_\varepsilon + SS_1)/(2\sigma_\varepsilon^2)$$

Hence, the conditional posterior for σ_ε^2 is an inverse gamma distribution with parameters $(d_\varepsilon + N)/2$ and $(d_\varepsilon v_\varepsilon + SS_1)/2$

Similarly, treating all parameters except σ_α^2 as fixed, the conditional posterior for σ_α^2 is obtained:

$$(\sigma_\alpha^2 | \mathbf{Y}, \theta, \sigma_\varepsilon^2) \propto (\sigma_\alpha^2)^{-(N d_\alpha + N/2 + 1)}$$
$$* \exp[-(d_\alpha v_\alpha + 2\Sigma_{AB}\{y_{ij} - \mu\}^2 + 2\Sigma_{BA}\{y_{ij} - \mu\}^2)/2\sigma_\alpha^2]$$

Hence, defining $SS_2 = 2\Sigma_{AB}(y_{ij} - \mu)^2 + 2\Sigma_{BA}(y_{ij} - \mu)^2$, the conditional posterior for σ_2^2 may be written:

$$(\sigma_2^2 | \mathbf{Y}, \boldsymbol{\theta}, \sigma_\varepsilon^2) \propto (\sigma_2^2)^{-((d_2+N)+1)} * \exp[-(d_2 v_2 + SS_2)/(2\sigma_2^2)]$$

making the conditional posterior for σ_2^2 an inverse gamma distribution with parameters $(d_2 + N)/2$ and $(d_2 v_2 + SS_2)/2$

The Gibbs sampler for σ_ε^2, σ_2^2, and $\boldsymbol{\theta}$ is thus specified by:

$$\sigma_\varepsilon^2 | \mathbf{Y}, \boldsymbol{\theta}, \sigma_2^2 = IG\{\tfrac{1}{2}(n+d_\varepsilon),\ \tfrac{1}{2}(SS_1 + d_\varepsilon v_\varepsilon)\} \quad (\sigma_\varepsilon^2 \leq \sigma_2^2)$$

$$\sigma_2^2 | \mathbf{Y}, \boldsymbol{\theta}, \sigma_\varepsilon^2 = IG\{\tfrac{1}{2}(n+d_2),\ \tfrac{1}{2}(SS_2 + d_2 v_2)\} \quad (\sigma_\varepsilon^2 \leq \sigma_2^2)$$

$$\boldsymbol{\theta} | \mathbf{Y}, \sigma_\varepsilon^2, \sigma_2^2 = N\{D(\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y} + \mathbf{C}^{-1}\mathbf{M}),\ D\}$$

Adding Carry-over to the Model

The addition of a first-order carry-over effect $\alpha = \alpha_A = -\alpha_B$, makes comparatively little difference to the above derivation. $\boldsymbol{\theta}$ is now a vector of four parameters μ, π, τ and α, and the design sub-matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ now have four columns, with

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix} \text{ and } \mathbf{X}_2 = \begin{pmatrix} 1 & 1 & -1 & 0 \\ 1 & -1 & 1 & -1 \end{pmatrix}$$

The variance-covariance matrix Σ for two observations on the same subject is unchanged, so that $\mathbf{X}'\Sigma^{-1}\mathbf{Y}$ has an extra element, to become a 4x1 vector, while $\mathbf{X}'\Sigma^{-1}\mathbf{X}$ has an additional row and column, becoming a 4x4 symmetric matrix. The additional element for $\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y}$ is

$\{(-\Sigma_{AB}y_{\bar{1}j} + \Sigma_{BA}y_{\bar{1}j})(12\sigma_d^2/\sigma_\varepsilon^2 + 1) + (\Sigma_{AB}y_{\bar{1}j} - \Sigma_{BA}y_{\bar{1}j})\}/\sigma_\varepsilon^2$

and the additional column (and row) of $\mathbf{X'S^{-1}X}$ is:

$(n_1 - n_2)/\sigma_\varepsilon^2$

$(n_2 - n_1)/\sigma_\varepsilon^2$

$-(n_1 + n_2)/\sigma_\varepsilon^2$

$\{(n_1 + n_2)\sigma_\varepsilon^2\}/(\sigma_\varepsilon^2\sigma_2^2)$

These changes affect the matrix **D**, and hence the mean and variance of the posterior distribution of θ, which is now a quadri-variate normal distribution, but the form of the mean and variance of this distribution is as before.

The means for $y_{\bar{1}j}$ and $y_{\bar{2}j}$ are now $\pi+\tau-\frac{1}{2}\alpha$ and $\pi-\tau+\frac{1}{2}\alpha$ respectively, and the means for $y_{\bar{1}j}$ and $y_{\bar{2}j}$ are $\mu+\frac{1}{2}\alpha$ and $\mu-\frac{1}{2}\alpha$ respectively, but the variances are unchanged. Hence it is necessary to re-define SS_1 and SS_2 as follows:

$SS_1 = 2\Sigma_{AB}\{y_{\bar{1}j} - (\pi+\tau-\frac{1}{2}\alpha)\}^2 + 2\Sigma_{BA}\{y_{\bar{2}j} - (\pi-\tau+\frac{1}{2}\alpha)\}^2$

$SS_2 = 2\Sigma_{AB}\{y_{\bar{1}j} - (\mu+\frac{1}{2}\alpha)\}^2 + 2\Sigma_{BA}\{y_{\bar{2}j} - (\mu-\frac{1}{2}\alpha)\}^2$

With these changes, the expressions given above for the conditional posterior distributions of σ_ε^2 and σ_2^2 still apply.

APPENDIX E

Derivation of Posterior Distributions for the Three-period Two-sequence Cross-over

Appendix E : Derivation of Posterior Distributions for the Three-period Two-sequence Cross-over

Posterior Distribution of θ

For the three-period design, it will still be assumed that the observations have a multivariate normal distribution ($\mathbf{Y} \sim N(\mathbf{X}\theta, \mathbf{S})$), and that the prior distribution of θ is also multivariate normal ($\theta \sim N(\mathbf{M}, \mathbf{C})$). Thus the expression given for the posterior distribution of θ given $\mathbf{X}$, $\mathbf{S}$, $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{Y}$ given in appendix D still applies. θ is now a vector of six parameters, it now being possible to include a sequence effect in the model, and two period effects are required. Thus $\theta' = (\mu, \gamma, \pi_1, \pi_2, \tau, \alpha)$, with $\gamma = \gamma_1 = -\gamma_2$ and $\pi_3 = -(\pi_1 + \pi_2)$. τ and α are as defined before for the two-period design.

The design sub-matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ are now 3x6 matrices:

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \end{pmatrix} \text{ and } \mathbf{X}_2 = \begin{pmatrix} 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix}$$

$$\text{and } \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_3^2 & \sigma_3^2 \\ \sigma_3^2 & \sigma_1^2 & \sigma_3^2 \\ \sigma_3^2 & \sigma_3^2 & \sigma_1^2 \end{pmatrix}$$

$$\text{so that } \Sigma^{-1} = (1/\sigma_2^2\sigma_3^2) \begin{pmatrix} \sigma_2^2 & -\sigma_3^2 & -\sigma_3^2 \\ -\sigma_3^2 & \sigma_2^2 & -\sigma_3^2 \\ -\sigma_3^2 & -\sigma_3^2 & \sigma_2^2 \end{pmatrix}$$

where $\sigma_1^2 = \sigma_u^2 + \sigma_e^2$, $\sigma_2^2 = 2\sigma_u^2 + \sigma_e^2$ & $\sigma_3^2 = 3\sigma_u^2 + \sigma_e^2$

Hence, for a subject in sequence 1:

$$\mathbf{X'\Sigma^{-1}Y} = \begin{pmatrix} (y_{ij1} + y_{ij2} + y_{ij3})/\sigma_3^2 \\ (y_{ij1} + y_{ij2} + y_{ij3})/\sigma_3^2 \\ (y_{ij1} - y_{ij3})/\sigma_1^2 \\ (y_{ij2} - y_{ij3})/\sigma_1^2 \\ (\sigma_1^2 y_{ij1} - \sigma_2^2(y_{ij2} + y_{ij3}))/(\sigma_1^2\sigma_3^2) \\ (y_{ij2} - y_{ij3})/\sigma_1^2 \end{pmatrix}$$

and $\mathbf{X'\Sigma^{-1}X} = 1/(\sigma_1^2\sigma_3^2)$

$$\begin{pmatrix} 3\sigma_1^2 & 3\sigma_1^2 & 0 & 0 & -\sigma_1^2 & 0 \\ 3\sigma_1^2 & 3\sigma_1^2 & 0 & 0 & -\sigma_1^2 & 0 \\ 0 & 0 & 2\sigma_3^2 & \sigma_3^2 & 2\sigma_3^2 & \sigma_3^2 \\ 0 & 0 & \sigma_3^2 & 2\sigma_3^2 & 0 & 2\sigma_3^2 \\ -\sigma_1^2 & -\sigma_1^2 & 2\sigma_3^2 & 0 & (3\sigma_1^2+8\sigma_u^2) & 0 \\ 0 & 0 & \sigma_3^2 & 2\sigma_3^2 & 0 & 2\sigma_3^2 \end{pmatrix}$$

and for a subject in sequence 2:

$$\mathbf{X'\Sigma^{-1}Y} = \begin{pmatrix} (y_{ij1} + y_{ij2} + y_{ij3})/\sigma_3^2 \\ -(y_{ij1} + y_{ij2} + y_{ij3})/\sigma_3^2 \\ (y_{ij1} - y_{ij3})/\sigma_1^2 \\ (y_{ij2} - y_{ij3})/\sigma_1^2 \\ (-\sigma_1^2 y_{ij1} + \sigma_2^2(y_{ij2} + y_{ij3}))/(\sigma_1^2\sigma_3^2) \\ (-y_{ij2} + y_{ij3})/\sigma_1^2 \end{pmatrix}$$

and $\mathbf{X'\Sigma^{-1}X} = 1/(\sigma_1^2\sigma_3^2)$

$$\begin{pmatrix} 3\sigma_1^2 & -3\sigma_1^2 & 0 & 0 & \sigma_1^2 & 0 \\ -3\sigma_1^2 & 3\sigma_1^2 & 0 & 0 & -\sigma_1^2 & 0 \\ 0 & 0 & 2\sigma_3^2 & \sigma_3^2 & -2\sigma_3^2 & -\sigma_3^2 \\ 0 & 0 & \sigma_3^2 & 2\sigma_3^2 & 0 & -2\sigma_3^2 \\ \sigma_1^2 & -\sigma_1^2 & -2\sigma_3^2 & 0 & (3\sigma_1^2+8\sigma_u^2) & 0 \\ 0 & 0 & -\sigma_3^2 & -2\sigma_3^2 & 0 & 2\sigma_3^2 \end{pmatrix}$$

so that $\mathbf{X'S^{-1}Y}$ will be:

$$\begin{pmatrix} \Sigma\Sigma(y_{ij1} + y_{ij2} + y_{ij3})/\sigma_s^2 \\ \{\Sigma(y_{1j1}+y_{1j2}+y_{1j3}) - \Sigma(y_{2j1}+y_{2j2}+y_{2j3})\}/\sigma_s^2 \\ \Sigma\Sigma(y_{ij1} - y_{ij3})/\sigma_e^2 \\ \Sigma\Sigma(y_{ij2} - y_{ij3})/\sigma_e^2 \\ \sigma_e^2\Sigma(y_{1j1}-y_{2j1}) - \sigma_{12}\Sigma(y_{1j2}+y_{1j3}-y_{2j2}-y_{2j3})/(\sigma_e^2\sigma_s^2) \\ \{\Sigma(y_{1j2}-y_{1j3}-y_{2j2}+y_{2j3})\}/\sigma_e^2 \end{pmatrix}$$

while $\mathbf{X'S^{-1}X}$ is:

$$\begin{pmatrix} 3(n_1+n_2) & 3(n_1-n_2) & 0 & 0 & -(n_1-n_2) & 0 \\ 3(n_1-n_2)/2 & 3(n_1+n_2) & 0 & 0 & -(n_1+n_2) & 0 \\ 0 & 0 & 2(n_1+n_2) & (n_1+n_2) & 2(n_1-n_2) & (n_1-n_2) \\ 0 & 0 & (n_1+n_2) & 2(n_1+n_2) & 0 & 2(n_1-n_2) \\ -(n_1-n_2)\sigma_e^2 & -(n_1+n_2)\sigma_e^2 & 2(n_1-n_2)\sigma_s^2 & 0 & (n_1+n_2)(3\sigma_e^2+8\sigma_s^2) & 0 \\ 0 & 0 & (n_1-n_2) & 2(n_1-n_2) & 0 & 2(n_1+n_2) \end{pmatrix}$$

all terms in rows 1 and 2 of the above matrix are divided by σ_s^2, all terms in rows 3, 4 & 6 by σ_e^2 and all terms in row 5 by $\sigma_e^2\sigma_s^2$.

These expressions allow the calculation of $\mathbf{D^{-1}}$ and hence $\mathbf{D}$, as before so that values of the conditional posterior for $\boldsymbol{\theta}$ can be simulated.

<u>Posterior Distribution of σ_s^2 & σ_e^2</u>

As before, new variables are defined which will be orthogonal. For the three-period design these variables are:

$$y_{1j}^{\dagger} = (2y_{ij1} - y_{ij2} - y_{ij3})/2\sqrt{3}$$

$$y_{2j}^{\dagger} = \tfrac{1}{2}(y_{ij2} - y_{ij3})$$

$$y\dot{i}_j = (y_{ij1} + y_{ij2} + y_{ij3})/\sqrt{6}$$

These three new variable have a tri-variate normal distribution, for a subject in the ABB sequence, the mean vector is:

$$\begin{pmatrix} (3\pi_1 + 4\tau)/2\sqrt{3} \\ ((\pi_1+2\pi_2) + 2\alpha)/2 \\ (3\mu + 3\gamma - \tau)/\sqrt{6} \end{pmatrix}$$

while for a subject in the BAA sequence, the mean vector is:

$$\begin{pmatrix} (3\pi_1 - 4\tau)/2\sqrt{3} \\ ((\pi_1+2\pi_2) - 2\alpha)/2 \\ (3\mu - 3\gamma + \tau)/\sqrt{6} \end{pmatrix}$$

for all subjects, the variance-covariance matrix is:

$$\frac{1}{2}\begin{pmatrix} \sigma_d^2 & 0 & 0 \\ 0 & \sigma_d^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

where $\sigma_3^2 = 3\sigma_a^2 + \sigma_d^2$

In terms of these new variables, the likelihood can be written:

$$\pi^{-(3N/2)}\sigma_d^{-2N}\sigma_3^{N}\exp[-\Sigma_{ABB}(y\dot{1}_j - (3\pi_1+4\tau)^2/\sigma_d^2]$$
$$*\exp[-\Sigma_{BAA}(y\dot{1}_j - (3\pi_1-4\tau))^2/\sigma_d^2]$$
$$*\exp[-\Sigma_{ABB}(y\bar{i}_j - (\pi_1+2\pi_2+2\alpha))^2/\sigma_d^2$$
$$*\exp[-\Sigma_{BAA}(y\bar{i}_j - (\pi_1+2\pi_2-2\alpha))^2/\sigma_d^2$$
$$*\exp[-\Sigma_{ABB}(y\dot{1}_j - (3\mu+3\gamma-\tau))^2/\sigma_3^2]$$
$$*\exp[-\Sigma_{BAA}(y\dot{1}_j - (3\mu-3\gamma+\tau))^2/\sigma_3^2]$$

The prior distribution for θ, σ_ε^2, and σ_β^2 is given by:

$$f(\theta, \sigma_\varepsilon^2, \sigma_\beta^2) = N(M, C)\, IG(\tfrac{1}{2}d_\varepsilon, \tfrac{1}{2}d_\varepsilon v_\varepsilon)\, IG(\tfrac{1}{2}d_\beta, \tfrac{1}{2}d_\beta v_\beta)$$

Hence, defining

$$SS_1 = 2\Sigma_{ABB}(y_{ij}^* - (3\pi_1+4\tau))^2 + 2\Sigma_{BAA}(y_{ij}^* - (3\pi_1-4\tau))^2$$

$$SS_2 = 2\Sigma_{ABB}(y_{ij}^- - (\pi_1+2\pi_2+2\alpha))^2 + 2\Sigma_{BAA}(y_{ij}^- - (\pi_1+2\pi_2-2\alpha))^2$$

$$SS_3 = 2\Sigma_{ABB}(y_{ij}^+ - (3\mu+3\gamma-\tau))^2 + 2\Sigma_{BAA}(y_{ij}^+ - (3\mu-3\gamma+\tau))^2$$

it is found that the Gibbs sampler for σ_ε^2, σ_β^2 and θ is:

$$\sigma_\varepsilon^2 | Y, \theta, \sigma_\beta^2 = IG(\tfrac{1}{2}(N+d_\varepsilon,\ \tfrac{1}{2}(SS_1/3 + SS_2 + d_\varepsilon v_\varepsilon)$$

$$\sigma_\beta^2 | Y, \theta, \sigma_\varepsilon^2 = IG(\tfrac{1}{2}(N+d_\beta,\ \tfrac{1}{2}(SS_3 + d_\beta v_\beta)$$

$$\theta | Y, \sigma_\varepsilon^2, \sigma_\beta^2 = N(D(X'S^{-1}Y+C^{-1}M), D)$$

subject to the restriction that $\sigma_\varepsilon^2 \leq \sigma_\beta^2$

Adding Second-order Carry-over to the Model

The addition of the second-order carry-over effect $\beta = \beta_A = -\beta_B$ to the model increases the size of θ and the design matrices X_1 and X_2, and hence $X'S^{-1}Y$ and $X'S^{-1}X$. The mean vectors of y^*, y^- and y^+ also change, necessitating a corresponding change in the definitions of SS_1, SS_2 and SS_3. These changes are similar to the changes caused by introducing first-order carry-over into the two-period design, and, as there, do not affect the form of the conditional posterior distributions.

APPENDIX F

Data Sets

Appendix F : Datasets

Hills & Armitage Enuresis Data

This data set is given in Hills & Armitage[1979], and has been widely used as an example of a data-set from the simple two-period cross-over. The data comes from a trial of a new drug for the treatment of enuresis, in which each of 29 patients were given the drug for a period of 14 days and the placebo for a separate period of 14 days, with the order of administration chosen randomly for each patient. The response recorded is the number of dry nights out of 14 for each treatment period.

Group 1 : drug, placebo			Group 2 : placebo, drug		
Patient Number	No of dry nights Period 1	Period 2	Patient Number	No of dry nights Period 1	Period 2
1	8	5	2	12	11
3	14	10	5	6	8
4	8	0	8	13	9
6	9	7	10	8	8
7	11	6	12	8	9
9	3	5	14	4	8
11	6	0	15	8	14
13	0	0	17	2	4
16	13	12	20	8	13
18	10	2	23	9	7

19	7	5	26	7	10
21	13	13	29	7	6
22	8	10			
24	7	7			
25	9	0			
27	10	6			
28	2	2			

Ebbutt's three-period data

In his paper on three-period cross-over designs, Ebbutt[1984] analysed the results of a trial which compared two treatments for hypertension. The trial had four sequence groups, ABB, BAA, ABA & BAB but only the data for the groups receiving ABB and BAA have been used in this thesis. Systolic and diastolic blood pressure was measured in each treatment period, and there was also a baseline measurement. Only the data for systolic blood pressure has been used, and these data are given below for the ABB and BAA sequence groups.

ABB Sequence

Baseline	Period 1	Period 2	Period 3
173	159	140	137
168	153	172	155
200	160	156	140
180	160	200	132

190	170	170	160
170	174	132	130
185	175	155	155
180	154	138	150
160	160	170	168
170	160	160	170
165	145	140	140
168	148	154	138
190	170	170	150
160	125	130	130
190	140	112	95
170	125	140	125
170	150	150	145
158	136	130	140
210	150	140	160
175	150	140	150
186	202	181	170
190	190	150	170

BAA Sequence

Baseline	Period 1	Period 2	Period 3
168	165	154	173
200	160	165	140
130	140	150	180
170	140	125	130
190	158	160	180

180	180	165	160
200	170	160	160
166	140	158	148
188	126	170	200
175	130	125	150
186	144	140	120
160	140	160	140
135	120	145	120
175	145	150	150
150	155	130	140
178	168	168	168
170	150	160	180
160	120	120	140
190	150	150	160
160	150	140	130
200	175	180	160
160	140	170	150
180	150	160	130
170	150	130	125
165	140	150	160
200	140	140	130
142	126	140	138

Biom. J. 31 (1989) 5, 589–598 Akademie-Verlag Berlin

Binary Responses and the Three-period Cross-over

G. H. MORREY

Leicester Polytechnic, School of Mathematical Sciences, England.

Summary

Problems with carry-over effects in the simple two-period cross-over have lead to interest in more complex cross-over designs. A method for analysing the optimum two-treatment three-period design with binary response variables is given by making a simple extension to Gart's logistic model. The method gives independent tests for, and estimates of the difference in treatment and first-order carry-over effects. An example of the analysis is given, using the loglinear models facility in GLIM.

Key words: Cross-over designs; Carry-over; Binary response variables; Logistic models; Log-linear models; GLIM; Clinical trials.

1. Introduction

The two-treatment, two-period crossover design has frequently been used in clinical trials to compare the action of treatments which relieve a condition. Sometimes the response observed is dichotomous, typically that the treatment is a 'success' or a 'failure'. Analysis of such data has been considered by several authors, with one of the best known expositions by GART [1969]. Gart uses a logistic model for the probability of success in each period with each sequence of treatments. The model uses the parameters

β_i = effect of ith subject
λ = period effect
δ = treatment effect

Note that β_i is the average effect of subject i while the period and treatment effect are regarded as deviations from this average with the usual restrictions $\lambda_1+\lambda_2=0$, $\delta_A+\delta_B=0$ so that $\lambda=\lambda_1=-\lambda_2$ and $\delta=\delta_A=-\delta_B$. Hence, for patient i in sequence one, who receives treatment A follows by treatment B:

$$P(\text{success in period 1 (treatment } A)) = \frac{e^{\beta_i+\lambda+\delta}}{1+e^{\beta_i+\lambda+\delta}}$$

$$P(\text{success in period 2 (treatment } B)) = \frac{e^{\beta_i-\lambda-\delta}}{1+e^{\beta_i-\lambda-\delta}}$$

Similar expressions can be obtained for subjects in sequence two, who receive treatment B in period one and treatment A in period two.

Writing 0 for 'failure' and 1 for 'success', there are four possible patterns of response: 00 01 10 11. The observed frequencies of these four patterns in sequence one will be denoted n_{00}, n_{01}, n_{10} and n_{11} respectively, while the observed frequencies in sequence two will be denoted n'_{00}, n'_{01}, n'_{10} and n'_{11}. It is helpful to form a score defined as response for period 1 minus response for period 2, which yields scores of $+1$,0 and -1. For sequence one, a score of $+1$ favours treatment A, while a score of -1 favours treatment B, with a score of 0 indicating no preference. For sequence two, scores of $+1$ favour treatment B, and scores of -1 favour treatment A.

The logistic model can be used to obtain expressions for the probability of any one of the four patterns of response, and when this is done it is found that apart from the normalising constant required to make the probabilities sum to one, only the probabilities relating to the patterns 01 and 10 contain the parameters δ and λ. Thus it seems reasonable to concentrate on those subjects who respond differently to the two treatments, obtaining a score of $+1$ or -1.

For a subject who experiences success on one treatment and failure on the other, the conditional probabilities of success on treatment A given only one success and success in period one given only one success are independent of the subject effect β_i, and these conditional probabilities can be used to give tests of $\lambda=0$ and $\delta=0$.

The two-period crossover design becomes difficult to analyse if there is any treatment-period interaction. For the two-period design such an interaction is usually represented by different carryover effects of the two treatments. Analyses of the two-period crossover with continuous data (Grizzle [1965], Hills & Armitage [1979]) test for such a difference in carry-over before testing for a difference between the treatments because the test for treatment differences is only valid if there is no difference in the carry-over from the two treatments. In Gart's logistic model there is no parameter to represent carry-over, and so no way to test for it.

2. A Three-period Crossover Design

Interest in two treatment crossover designs with more than two periods was stimulated by a paper which reviewed the properties of such designs (Kershner & Federer [1981]). This suggested that the three-period design in which two treatment groups receive the treatment sequences ABB and BAA respectively is particularly efficient in giving estimates of the treatment and carry-over effects. It has since been confirmed that this design is the optimal three-period design when carry-over is present (Laska, Meisner & Kushner [1983]).

Analysis of this design with a continuous response variable has been dealt with elsewhere (Morrey [1984], Ebbutt [1984]) and the purpose of this paper is to deal

with the analysis of the design with dichotomous responses. Following Gart, a logistic model is adopted, but this now includes two parameters for periods (λ_1, λ_2) and a term for the carry-over effect of a treatment from one period to the next, represented by the parameter c.

These period and carry-over effects are again regarded as deviations from the average subject effect with restrictions $\lambda_1+\lambda_2+\lambda_3=0$ i.e. $\lambda_3=-\lambda_1-\lambda_2$ and $c_A+c_B=0$ i.e. $c=c_A=-c_B$.

First-order carry-over takes up one of the two degrees of freedom for treatment $\times$period interaction in the three-period design, so there may still be aspects of treatment$\times$period interaction not allowed for in the model. However, it is hoped that first-order carry-over would form a major part or any treatment$\times$period interaction, and that any remaining interaction would be comparatively uninteresting. It would certainly be difficult to interpret.

Using this model, the probabilities of success in each of the three periods for subject i in sequence one are:

$$p\text{ (success in period 1 (treat } A)) = \frac{e^{\beta_i+\lambda_1+\delta}}{1+e^{\beta_i+\lambda_1+\delta}}$$

$$p\text{ (success in period 2 (treat } B)) = \frac{e^{\beta_i+\lambda_2-\delta+c}}{1+e^{\beta_i+\lambda_2-\delta+c}}$$

$$p\text{ (success in period 3 (treat } B)) = \frac{e^{\beta_i-\lambda_1-\lambda_2-\delta-c}}{1+e^{\beta_1-\lambda_1-\lambda_2-\delta-c}}$$

Similar expressions can be obtained for a subject in sequence two.

There are now eight possible patterns: 000, 001, 010, 011, 100, 101, 110, 111; as before, the observed frequencies of these patterns will be labelled n_{000}, n_{001} etc. for sequence one and n'_{000}, n'_{001} etc. for sequence two. It is again useful to define a score, and consideration of the analysis for the design with continuous response variables suggests the combination twice period 1 minus period 2 minus period 3 might be useful. Such a combination yields scores of -2, -1, 0, 1 or 2, and produces a logical ordering of the eight patterns, as shown in table 1. The model can be used to give the probability of any of these patterns, and the probabilities for subject i in sequence one are also tabulated in table 1. Corresponding probabilities for a subject in sequence two would differ in the sign of the terms in δ and c.

It can be seen from table 1 that apart from the normalising constant, the term for carry-over occurs only in the patterns giving a score of -1 or $+1$. Thus, by concentrating on subjects who obtain a score of -1, or on those who score $+1$, it may be possible to obtain a test for carry-over. Note that the patterns giving scores of -1 and $+1$ are of two basic types: type R have a success in period two (010 or 110), while type S have a success in period three (001 or 101).

(1) for sequence 1. p (type R pattern/score $=-1$) $= \dfrac{e^{\lambda_2+c}}{e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c}}$

(2) for sequence 2. p (type R pattern/score $=-1$) $= \dfrac{e^{\lambda_2-c}}{e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c}}$

Table 1

Scores for each pattern, and probabilities for subject i in sequence one.
Note: $k=[1+\exp(\beta_i+\lambda_1+\delta)][1+\exp(\beta_i+\lambda_2-\delta+c)][1+\exp(\beta_i-\lambda_1-\lambda_2-\delta-c)]$

score	pattern	observed frequency seq 1	observed frequency seq 2	implication for sequence 1	implication for sequence 2	probability for subject i in sequence 1.
−2	011	n_{011}	n'_{011}	most favours B	most favours A	$\frac{1}{k}\exp(2\beta_i-\lambda_1-2\delta)$
1	010	n_{010}	n'_{010}	favours B	favours A	$\frac{1}{k}\exp(\beta_i+\lambda_2-\delta+c)$
	001	n_{001}	n'_{001}			$\frac{1}{k}\exp(\beta_i-\lambda_1-\lambda_2-\delta-c)$
0	000	n_{000}	n'_{000}	no preference	no preference	$\frac{1}{k}$
	111	n_{111}	n'_{111}			$\frac{1}{k}\exp(3\beta_i-\delta)$
1	110	n_{110}	n'_{110}	favours A	favours B	$\frac{1}{k}\exp(2\beta_i+\lambda_1+\lambda_2+c)$
	101	n_{101}	n'_{101}			$\frac{1}{k}\exp(2\beta_i-\lambda_2-c)$
2	100	n_{100}	n'_{100}	most favours A	most favours B	$\frac{1}{k}\exp(\beta_i+\lambda_1+\delta)$

Clearly, the above expressions are identical if there is no difference in the carry-over effect of the two treatments, making $c=0$. Thus a test for carry-over can be obtained by using the 2×2 table:

	sequence 1	sequence 2
pattern R	n_{010}	n'_{010}
type S	n_{001}	n'_{001}

Alternatively, considering only subjects who score $+1$:

(3) for sequence 1: p (pattern type R/score $=+1$) $=\dfrac{e^{\lambda_2+c}}{e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c}}$

(4) for sequence 2: p (pattern type R/score $=+1$) $=\dfrac{e^{\lambda_2-c}}{e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c}}$

These conditional probabilities are identical to the previous two, providing not only an alternative test for carry-over, but also presenting an opportunity of testing whether the underlying model is reasonable. The $2\times2\times2$ contingency table:

score = −1	sequence 1	sequence 2
pattern R	n_{010}	n'_{010}
type S	n_{001}	n'_{001}

score = +1	sequence 1	sequence 2
pattern R	n_{110}	n'_{110}
type S	n_{101}	n'_{101}

could be analysed using a log-linear model in which terms for sequences, scores and patterns are fitted. Since conditional probabilities (1) and (3), and (2) and (4) above are identical, it is clear that the model is specifying that there will be no three-factor interaction or interaction between patterns and scores. Testing for these interactions is thus a test of the underlying model. If $c = 0$ the conditional probabilities (1), (2), (3) and (4) are all equal, and there is no interaction between patterns and sequences, so testing for this interaction gives a test of carry-over.

It is possible to obtain an estimate of c by equating the expressions for the probabilities in equations (1) to (4) with the corresponding observed relative frequencies. Since the expressions in (1) and (3) and in (2) and (4) are identical the two relative frequencies should be combined in some way. The ideal would be to weight them by the inverse of their standard errors, but this is not possible as the standard errors contain the parameters being estimated, so a more ad hoc method must be used. The two obvious possibilities are to average the two relative frequencies, or to add the numerators and denominators. The latter method results in a simpler expression for the estimate of c, and seems the more preferable, giving

$$\frac{n_{010}+n_{110}}{n_{010}+n_{001}+n_{110}+n_{101}} \quad \text{as an estimate of} \quad \frac{e^{\lambda_2+c}}{e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c}}$$

and

$$\frac{n'_{010}+n'_{110}}{n'_{010}+n'_{001}+n'_{110}+n'_{101}} \quad \text{as an estimate of} \quad \frac{e^{\lambda_2-c}}{e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c}}$$

writing

$$L=\frac{e^{\lambda_2+c}}{e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c}} \quad \text{and} \quad M=\frac{e^{\lambda_2-c}}{e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c}}$$

it is easy to see that $\dfrac{L}{1-L}\dfrac{1-M}{M}=e^{4c}$ so that the required estimate of c is

$$\hat{c}=\frac{1}{4}\ln\left\{\frac{n_{010}+n_{110}}{n_{001}+n_{101}}\cdot\frac{n'_{001}+n'_{101}}{n'_{010}+n'_{110}}\right\}$$

From table 1 it can be seen that apart from the normalising constant, the term for treatment effect in the model appears in the probabilities for the patterns in which there is just one success: 010, 001, 100; and in the pattern giving a score of −2 (011). The term also appears in the probability for the 111 pattern, but this information about δ is difficult to retrieve. Concentrating on those patterns with just one success, we have the conditional probabilitities:

for sequence 1 : p (success in period 1 | 1 success) $=\dfrac{e^{\lambda_1+\delta}}{e^{\lambda_1+\delta}+e^{\lambda_2-\delta+c}+e^{-\lambda_1-\lambda_2-\delta-c}}$

for sequence 2 : p (success in period 1 | 1 success) $=\dfrac{e^{\lambda_1-\delta}}{e^{\lambda_1-\delta}+e^{\lambda_2+\delta-c}+e^{-\lambda_1-\lambda_2+\delta+c}}$

Clearly these conditional probabilities are equal if δ and c are both zero so that testing the equality of these would test whether $\delta = 0$ if the previous test showed that $c = 0$. However, it would be preferable to obtain a test of $\delta = 0$ allowing for the effect of first-order carry-over.

The maximum likelihood estimators of the above conditional probabilities will be the observed relative frequencies

$$\frac{n_{100}}{n_{100}+n_{010}+n_{001}} \quad \text{and} \quad \frac{n'_{100}}{n'_{100}+n'_{010}+n'_{001}}$$

if these estimators are equated with the expressions for the conditional probabilities that they estimate, an algebraic simplification can be made yielding:

$$\text{for sequence 1: } e^{-\delta}\,(e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c}) = \frac{n_{010}+n_{001}}{n_{100}}\,e^{\lambda_1+\delta}$$

$$\text{for sequence 2: } e^{\delta}\,(e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c}) = \frac{n'_{010}+n'_{001}}{n'_{100}}\,e^{\lambda_1-\delta}$$

dividing these gives

$$(5) \qquad e^{-2\delta}\left\{\frac{e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c}}{e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c}}\right\} = \left(\frac{n_{010}+n_{001}}{n'_{010}+n'_{001}}\right)\frac{n'_{100}}{n_{100}}\,e^{2\delta}$$

Similarly, if the patterns giving exactly two successes, or just one failure, are considered, we can obtain the conditional probabilities:

$$\text{for sequence 1: } p\text{ (failure in period 1 | 1 failure)} = \frac{e^{-\lambda_1-2\delta}}{e^{-\lambda_1-2\delta}+e^{\lambda_1+\lambda_2+c}+e^{-\lambda_1-c}}$$

$$\text{for sequence 2: } p\text{ (failure in period 1 | 1 failure)} = \frac{e^{-\lambda_1+2\delta}}{e^{-\lambda_1+2\delta}+e^{\lambda_1+\lambda_2-c}+e^{-\lambda_2+c}}$$

Again, equating these to the observed relative frequencies which estimate them and simplifying gives:

$$\text{for sequence 1: } \frac{n_{110}+n_{101}}{n_{011}}\,e^{-\lambda_1-2\delta} = e^{\lambda_1}\,(e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c})$$

$$\text{for sequence 2: } \frac{n'_{110}+n'_{101}}{n'_{011}}\,e^{-\lambda_1+2\delta} = e^{\lambda_1}(e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c})$$

dividing these gives:

$$(6) \qquad e^{-4\delta}\left(\frac{n_{110}+n_{101}}{n'_{110}+n'_{101}}\right)\frac{n'_{011}}{n_{011}} = \frac{e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c}}{e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c}}$$

If $\delta = 0$ the L.H.S. of equation (5) is equal to the R.H.S. of equation (6), implying that:

$$\frac{(n_{010}+n_{001})\,n'_{100}}{(n'_{010}+n'_{001})\,n_{100}} = \frac{(n_{110}+n_{101})\,n'_{011}}{(n'_{110}+n'_{101})\,n_{011}}$$

The equality of these ratios is therefore a test of $\delta = 0$. This can be tested by testing the three-way interaction in the $2 \times 2 \times 2$ table:

1 success	sequence 1	2
in per no	$n_{010}+n_{001}$	$n'_{010}+n'_{001}$
1? yes	n_{100}	n'_{100}

1 failure	sequence 1	2
in per no	$n_{101}+n_{110}$	$n'_{101}+n'_{110}$
1? yes	n_{011}	n'_{011}

Note also that the period $\times$ sequence interaction is also a test of $\delta = 0$ if there is no first-order carry-over effect i.e. $c = 0$.

An estimate of δ can be obtained from the expressions (5) and (6) above. From equation (5)

$$e^{4\delta} = \left\{\frac{e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c}}{e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c}}\right\} \div \left\{\frac{n_{010}+n_{001}}{n'_{010}+n'_{001}} \cdot \frac{n'_{100}}{n_{100}}\right\}$$

while equation (6) gives

$$e^{4\delta} = \left\{\frac{e^{\lambda_2+c}+e^{-\lambda_1-\lambda_2-c}}{e^{\lambda_2-c}+e^{-\lambda_1-\lambda_2+c}}\right\} \div \left\{\frac{n_{101}+n_{101}}{n'_{110}+n'_{101}} \cdot \frac{n'_{011}}{n_{011}}\right\}$$

dividing these gives

$$\delta = \frac{1}{8} \ln \left\{\frac{n_{110}+n_{101}}{n'_{110}+n'_{101}} \cdot \frac{n'_{011}}{n_{011}}\right\} \div \left\{\frac{n_{010}+n_{001}}{n'_{010}+n'_{001}} \cdot \frac{n'_{100}}{n_{100}}\right\}$$

3. Example

The ease with which these tests can be applied can be seen by considering their use with a set of data. The data used here are the expected frequencies, rounded to the nearest integer, obtained by using the linear model for the logits given above with parameter values $\lambda_1 = 0.2$, $\lambda_2 = 0.3$, $\delta = 0.5$ and $c = 0.6$, and assuming $N = N' = 20$.

response	000	100	010	001	110	101	011	111
observed seq. 1 (ABB)	5	4	3	0	5	1	1	1
freq. seq. 2 (BAA)	4	1	2	3	1	2	4	3

Note that if the usual two-period crossover had been undertaken the following responses would have been observed:

response	00	01	10	11
observed seq. 1 (AB)	5	4	5	6
freq. seq. 2 (BA)	7	6	3	4

Gart's tests would then give:

for treatment effects

	seq. 1 (AB)	seq. 2 (BA)
success per. 1	5	3
in per. 2	4	6

$\chi^2 = 0.9$ (not sig)

for period effect

	seq. 1 (AB)	seq. 2 (BA)
success treat A	5	6
on treat B	4	3

$\chi^2 = 0.23$ (not sig).

It is of particular interest that Gart's test fails to detect the difference in treatments.

Using the proposed new tests with the data from all three periods, the first $2 \times 2 \times 2$ contingency table to be considered is

score $= -1$	seq. 1	seq. 2
pattern R	3	2
S	0	3

score $= +1$	seq. 1	seq. 2
pattern R	5	1
S	1	2

This was analysed using a log-linear model fitted by GLIM with the three factors sequence (Q), score (S) and Pattern (P). This gives the following scaled deviances:

3 – factor interaction $= 0.56453$
pattern x score interaction $= 0.3435$
i.e. 0.90805 on 2 d.f. (not sig)

Recall that the non-significance of these two interactions is a test of the model, which in this case is known to be appropriate as the data was generated from it. The test of carryover is based on the pattern $\times$ sequence interaction which has a scaled deviance of 5.21 on 1 d.f. (sig at 5%). Thus the test correctly detects the presence of carryover.

The estimate of the carryover effect $\hat{c}$ is

$$\frac{1}{4} \ln \left\{ \frac{3+5}{0+1} \cdot \frac{3+2}{2+1} \right\} = 0.6476$$

This may seem a rather poor estimate of the true value of 0.6, but it must be remembered that the true expected frequencies have been rounded to the nearest integer, which causes the inaccuracy.

The second $2 \times 2 \times 2$ contingency table is as follows:

	sequence	
1 success	1	2
in per no	3	5
1? yes	4	1

	sequence	
1 failure	1	2
in per no	6	3
1? yes	1	4

This has again been analysed by fitting a log-linear model using GLIM in which the three factors sequence (Q), success/failure (S), and period (P) are used.

The three-factor interaction is a test of treatments and the scaled deviance for this is 5.279 on 1 d.f. (sig. at 5%) so that the test again correctly detects the presence of a treatment effect.

The estimate of the treatment effect δ is

$$\frac{1}{\delta} \ln \left\{ \frac{(5+1)}{(1+2)} \cdot \frac{4}{1} \right\} \div \left\{ \frac{(3+0)}{(2+3)} \cdot \frac{1}{4} \right\} = 0.497$$

gratifyingly near to the true value of 0.5 in spite of the innaccuracies introduced by rounding the expected frequencies.

4. Discussion

In this paper, Gart's logistic model for the two-period crossover has been extended to the three-period crossover. Gart's model has individual subject parameters β_t, which will accommodate any dependency between the observations on the same subject, whereas a recent paper by KENWARD & JONES [1987] uses a parameterisation with specific terms in the model to describe the within-subject dependency structure. This latter type of model could be used for the three-period design, but it was felt that Gart's parameterisation was more intuitive.

The drawback with Gart's model for the two-period crossover and to some extent the design itself is that no test of carry-over is possible, although it is well-known that any test of treatments is invalid if difference in carry-over exists. HILLS and ARMITAGE [1979] do give a test for carry-over for the two-period crossover with binary data, and although no formal justification of the test is given, KENWARD & JONES [1987] have subsequently shown that it does test for difference in carry-over in the two groups. HILLS & ARMITAGE reason that carry-over will tend to alter the average number of successes in the two groups, and hence the relative proportions of subjects giving responses 00 and 11. This is also the basis for the corresponding test with continuous response variables, which is known to lack power (BROWN [1980]), so that the test for binary data is also likely to be weak. More important is the fact that the test for carry-over is essentially a pre-test which must give a nonsignificant result if the test for treatments is to be valid. The three-period crossover avoids this difficulty, allowing a test for treatments which is valid even if there is a difference in first-order carry-over, and hence is to be preferred. It is hoped that, by showing that tests for the three-period design can be simply performed using log-linear models, this paper will encourage the use of the three-period design.

References

BROWN, B. W. Jr., 1980: The crossover experiment for clinical trials. Biometrics **36**, 69–79.

EBBUTT, A. F., 1984: Three-period crossover design for two treatments. Biometrics **40**, 219–224.

GART, J. J., 1969: An exact test for comparing matched proportions in crossover designs. Biometrika **56**, 75–80.

GRIZZLE, J. E., 1965: The two-period change-over design and its use in clinical trials. Biometrics **21**, 467–480. Corrigenda (1974) Biometrics **30**, 727.

HILLS, M. & ARMITAGE, P., 1979: The two-period crossover clinical trial. Brit. J. Clin. Pharmac., 8, 7–20.

KENWARD, M. G. & JONES, B., 1987: A log-linear model for binary and crossover data. Appl. Statist. **36**, 192–204.

KERSHNER, R. P. & FEDERER, W. T., 1981: Two-treatment crossover designs for estimating a variety of effects. JASA **76**, 612–619.

LASKA, E. MEISNER, M. & KUSHNER, H. B., 1983: Optimal crossover designs in the presence of carryover effects. Biometrics **39**, 1087–1091.

MORREY, G. H., 1984: A simple alternative to the two-period crossover design. Leicester Statistics Technical Report No. 2.

Received May 1988

G. H. MORREY
Leicester Polytechnic
Leicester LE1 9BH, England